CHANGE DETECTION WITH CLIMATE DATA APPLICATION

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA

BY **YING LU**

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Snigdhansu Chatterjee

August, 2015

Acknowledgments

This dissertation is dedicated to Doctor of Philosophy in Statistics at the University of Minnesota-Twin Cities. I am very privileged to have the opportunity to work with Professor Snigdhansu Chatterjee. He has shown me how to do a good scientific research by inspiring my ideas and guiding me to the wonderful world of statistics. He has given me enormous support and encouragement, especially when I was frustrated. The training and research experience I have received from him has been invaluable and will benefit my entire career.

I would also like to thank Professor Glen Meeden, Lan Wang, Vipin Kumar and Arindam Banerjee. They have spent huge amount of time in reviewing my thesis, and have been so generous in providing me with valuable feedbacks. In addition, I would like to thank Stefan Liess for his comments on some of the intriguing topics in the climate science community.

I also wish to extend my thanks to Professor Zou Hui, Yuhong Yang, Galin Jones and all the professors who give me support during my PhD studies.

Finally to my wife Jenen Tan, and my parents Tangyong Lu and Zhilan Wei, and many others, whose encouragement made this work possible, I dedicate this thesis.

This research is partially supported by NSF grants IIS-1029711 (Collaborative Research: Mining Climate and Ecosystem Data) and SES-0851705 (Collaborative Research: Computation-driven small area inference with applications).

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Hurricanes are some of the most destructive and costliest natural disasters of the world. On average at least one hurricane strikes Florida every two years. Eight of the ten most expensive hurricanes ever to make landfall in US history have had at least some effect on Florida, causing in excess of 60 billion in insured losses, see [100]. As populations continue to increase in Florida coastal regions where the threat is highest, so does the possibility for even greater destruction. Therefore, it is crucial to study hurricane behavior for better understanding and possible mitigation to the loss of property and lives.

One important way in which hurricane patterns may be present in data related to many physical and natural phenomena is by structural breaks and changes. Generally, elicitation of the time and nature of such breaks with statistical guarantees involves change detection techniques like the *cumulative sum* (CUSUM), or the *exponentially weighted moving average* (EWMA).

The standard framework for applying such change detection techniques requires assuming that the order in which the sampled observations arrive is known, with the question of interest being whether the data generating process has remained stable over time. Typically, the observations are assumed to follow a known Gaussian distribution, and are monitored for a potential change to a different, but still known, Gaussian distribution. Statistical guarantees are typically expressed in terms of expected *run length*, *i.e.*, how long it takes on average for a true change to be detected, when there is a control for the expected length of time before false signaling occurs. These normality-based sequential monitoring and stability detection techniques originated from industrial process control [112], although they have far-ranging applications nowadays. Examples of such applications are in health care monitoring [136], detection of genetic mutation [82], credit card and financial fraud detection [9], insider trading in stock markets [105], and detection of jamming attacks in wireless networks [21].

Note that in many modern applications, the assumption of normality is not tenable. The statistical problem we propose is described as follows: Suppose we observe a series of random variables $\{X_n\}_{n=1}^{\infty}$ with $X_1, \dots, X_\tau$ are identically and independently distributed from $F(\theta)$, whereas $X_{\tau+1}, \dots, X_n$.. are identically and independently distributed from $F(\tilde{\theta})$. The distribution $F(\cdot)$, as well as the parameters θ and $\tilde{\theta}$ are known, except for the change time τ, which is fixed but unknown, $0 \leq \tau \leq \infty$. $\tau = \infty$ means there is no change in the series, $\tau = 0$ means there is a change from the beginning of the series, and $0 < \tau < \infty$ means there is a change at some point in the process.

In order to generalize the scope of statistical change detection tools, we propose a variant of the sequential industrial monitoring framework, by considering the stability of the data generation process as a problem of detecting the *time of the distributional change*. That is, we conduct a hypothesis test, and under the null hypothesis, the data generation process remains stable through the entire sampling time $t = 1, \dots, n$. Under the alternative hypothesis, the distribution of the individual observations remains stable up to an unknown point of time $\tau \leq n$ and then it changes to another distribution. With this hypothesis testing framework, we are in a position to (a) consider models with zero, one or more change points in the same statistical framework, (b) quantify uncertainty associated with any potential result using standard concepts of hypothesis tests like size, power, level of significance, or properties of the run length, (c) extend the scope of the study beyond the traditional frameworks where the data either arrives sequentially, or there are sufficient observations before and after each change point. We may consider problems where some parameters are known for some duration of the process, while others are estimated. The sequential process monitoring statistics like CUSUM are obtained as a special case, so there is no loss of generality in using the hypothesis testing approach proposed here. We call the proposed testing procedure the *exponential family CUSUM* (or EF-CUSUM in short), while the statistic obtained under the Gaussian framework is called normal-CUSUM or Gaussian-CUSUM.

In our formulation, a change detection problem is one of hypothesis testing, where $H_0 : \tau \geq n \; VS \; H_1 : 0 \leq \tau < n$. To solve the hypothesis testing problem, we construct

$$\Lambda = \frac{\max_{0 \leq \tau < n} L(X_1, \dots, X_n; \theta')}{\max_{\tau \geq n} L(X_1, \dots, X_n; \theta)}.$$

The function $L(X_1, \dots, X_n; \theta)$ is the joint likelihood function of $X_1, \dots, X_n$ when the underlying parameter is θ. The decision rule is in the form of $\Lambda \geq L$, where L is usually determined by the Type I error. Intuitively, when Λ becomes large, we tend to conclude a change in the process, while when Λ is small, we do not have sufficient evidence for a change.

The ideal algorithm is to signal a change as quickly as possible, but not to signal a change when no change occurs. In reality, it is impossible to achieve both goals. We therefore pursue an algorithm that signals a change as quickly as possible while controlling early signals under a pre-specified amount when no change occurs. One popular measure of performance is the *average run length*. To formally introduce it, a *run length* N is a stopping time, *i.e.* N is measurable with respect to the σ-algebra generated by $X_1, \ldots, X_N$. Let P_m be the distribution of $X_1, X_2...$ under which X_m is the first term with distribution F_1. Let E_m be the expectation under P_m. This was quantified by [91] as the problem of minimizing

$$ARL_1 = \big(\bar{E}_1 N =\big) \sup_{m \geq 1} ess \sup E_m[(N - m + 1)^+ | x_1, \ldots, x_{m-1}],$$

while keeping $ARL_0 = E_\infty N \geq \gamma$ for pre-defined $\gamma > 0$. Note that ARL_1 is a mini-max formulation of the expected run length after the change has taken place. In this framework, such probability ratio testing procedure was optimal, see [107].

In Chapter 2, we discuss the change detection topic in the context of exponential family and regression models including generalized linear models such as logistic regression and log-linear regression. We present several mathematical results concerning the different kinds of CUSUM statistics that may result, depending on the probabilistic structure under consideration, and whether certain parameters are estimated or assumed known. A natural question here is on the performance of the normality-based CUSUM statistic, when the probability models do not satisfy the Gaussian assumptions. We study this issue, and present mathematical results, simulation studies and discussions about when and how the Gaussian-CUSUM may yield high quality results.

Simulation studies show that in most situations, EF-CUSUM method performs better than Gaussian-CUSUM. The EF-CUSUM has a shorter average run length, smaller variation of run length and shorter maximum run length compared with Gaussian-CUSUM. Moreover, smaller shifts can be detected more quickly by EF-CUSUM than by Gaussian-CUSUM, which is a big advantage of using EF-CUSUM. Under some circumstances the Gaussian-CUSUM approximates the EF-CUSUM well. It is also important to note that whether the change point τ is at the beginning, in the middle or at the end, the EF-CUSUM generally outperforms the Gaussian-CUSUM, so the unknown parameter τ plays little role in our analysis. Finally, in the case of a large parameter shift, the exponential family CUSUM and the Gaussian-CUSUM perform similarly. This is not unusual, and even visual and *ad hoc* techniques suffice for many cases of large changes.

We also extend our study to the regression parameter change context in the generalized linear model. We would like to know if the nonlinear relationship between the response and covariates remains unchanged over time. The corresponding EF-CUSUM scheme, which leverages the link function and the underlying probability structure, is developed and discussed.

Our case study for illustrating our instability and change detection techniques is based on the Atlantic hurricane data. There are several studies recently on whether, and how, the properties of these storms have changed with climate change, see for example [124]. Apart from being of current interest, the presence of some amount of evidence for change in the literature is helpful for evaluating whether our proposed methods can detect known instabilities.

Throughout the thesis from Chapter 2 to 5, we will study the Atlantic hurricane data and revisit it a few times. Atlantic hurricanes generally originate starting from early summer (May) till autumn (November). We use 1951-2008 Atlantic hurricane data from http://weather.unisys.com/hurricane/atlantic/, which imports data from the National Weather Service via the NOAAPORT satellite data service. National Oceanic and Atmospheric Administration(NOAA) tracks and records each hurricane every six hours from birth to death about the date, time, id, location, maximum sustained wind speeds and central pressure. In addition, we collect climate variables such as North Atlantic Oscillation(NAO), Southern Oscillation Index(SOI), Sea Surface Temperature(SST), Carbon Dioxide (CO2) and El Nino/La Nina data from other sources.

In the hurricane analysis in Chapter 2, We study the yearly number of such storms, as well as the joint relationship between pressure and wind speeds. We detect changes compatible with known facts. Interestingly, we find that although wind speed and central pressure values of Atlantic tropical storms have changed, they have changed in-sync, that is, their mutual relationship has remained stable over time. This lends credence that our methodology might be able to detect true changes and discard false signals well, since large scale energy balance relationships (as that between pressure and wind speed) are not expected to change.

Our second topic of interest is change detection in extreme events, which are usually characterized by the family of generalized extreme value distribution (GEV). The GEV distribution is the limit distribution of properly normalized maxima of a sequence of random variables that are independent and identically distributed. This family of distribution can be classified into three categories: Gumbel, Fréchet and reversed Weibull distribution, which are known as type I, type II and type III extreme value distributions. The study of extreme events is very important in the field of engineering, finance, insurance, climate and so on. In engineering, dams and buildings are constructed to prevent the rare, extreme natural disasters such as flood and earthquake. In finance, bankers implement rigorous risk management strategy to prevent loss leading to bankruptcy. For insurance and reinsurance companies, they are most worried about extreme weather conditions, and given damages are correlated, their premium calculation should carefully incorporate such scenarios. Many more applications can be found in the following: [83], [25], [24], [55], [79], [83], [111], [120], [53], [135], [13] and [4].

In Chapter 3, we first follow similar approach described in Chapter 2 and develop the GEV-CUSUM procedure, including Gumbel, Fréchet, Weibull and Generalized Pareto distribution. One drastic difference between the GEV-CUSUM and EF-CUSUM is that the support of the GEV distribution largely depends on the parameter, while for the exponential family, sufficient statistic is easily obtained via the factorization theorem, thus entirely independent of the underlying parameter. Therefore, no matter whether the underlying parameter of the GEV distribution is known or not, the corresponding GEV-CUSUM procedure requires many more scenarios to be taken care of because of the nature of the distribution.

Secondly, we study and compare the performance of normal-CUSUM and our proposed GEV-CUSUM through simulation study. Besides the average run length mentioned earlier, we also propose the *p-value approach.* The idea is that under the circumstrance when there is no change of distribution, the corresponding log likelihood ratio statistic forms a distribution, which we call the null distribution. When there is a change in the process, the log likelihood ratio statistic has a different distribution. We compute the tail probability for that statistic against the null, and call it p-value. Finally by collecting the 25th, 50th, 75th and 90th percentile of those p-values for each possible change point for normal-CUSUM and GEV-CUSUM, we are able to compare between different procedures: the smaller the p-value, the more powerful the procedure is.

When we compare the normal-CUSUM versus GEV-CUSUM procedure, we need to impose the constraint that the mean and variance for the two distributions being equal for fair comparison. Simulation results show that the GEV-CUSUM procedure dominates the normal-CUSUM procedure in the sense that it gives shorter average run length and smaller 90, 75, 50 and 25 percentiles of p-values after a change occurs, irrespective of where the change point τ is located.

In the last section of Chapter 3, we conduct an analysis on the maximum sustained wind speeds. Papers such as [33], [46], [43] modeled extreme wind speeds as generalized Pareto distribution or Weibull distribution, and studied the upper quantiles of wind speed. Furthermore, papers such as [70] related climate covariates with the scale and shape parameter for the Weibull distribution for the maximum sustained wind speed, and used simulation to determine the annual exceedance probability of the wind speeds at all levels. In Chapter 3, we focus on leveraging our GEV based approach to study the changing behavior of the maximum sustained wind speeds.

In Chapter 4, we generalize the GEV-CUSUM approach by joint modeling the order statistic. Order statistic is a more comprehensive way of describing extreme events than maxima (minima) as it contains more information with inherent extreme patterns. As a result, the test power is stronger than if only the maxima (minima) is considered. In other words, if there is truly a change in the distribution for GEV distribution, using order statistic leverages the data power and detects the change more effectively. Another advantage is that the analysis would be more robust. More data points within each block enter the analysis, which largely mitigates the side effect of outliers or even just an error in data entry. There is trade off, however, in determining the number of order statistic to use. Using a large number of order statistic destroys the asymptotic behavior, making the generalized extreme value distribution a poor approximation. Using a small number of order statistic does not leverage the data set enough, leading to loss of efficiency. Moreover, this method yields more computational complexity, and require more advanced techniques including reject sampling and constrained optimization, since there is no easy method to deal with such a situation.

At the end of Chapter 4, we revisit the example of Atlantic hurricane and discuss the maximum sustained wind speeds with the technique of order statistic. Papers such as [22] mentions order statistic for sea level study. We adopt our newly proposed r-order statistic method to study the changing behavior of the maximum sustained wind speeds. Empirical study shows evidence of change in the late 1960s. It further demonstrates that as r increases, the p-value seems to exhibit a decreasing trend while the estimated change point does not fluctuate much. This indicates that when we include order statistic, the evidence of change becomes stronger, and we will conclude a change which otherwise we won't if only the maxima is used. The downside of including too many order statistic is shown in the case of $r = 5$. In this case, since the asymptotic GEV approximation may be poor, the estimated change point will be inconsistent with what we obtain for smaller r.

Finally, in Chapter 5, we move from the change detection topic to the data mining field. There, we spend the whole chapter on the hurricane trajectory prediction problem. Researchers have been working on various forecasting models for several decades, and among the best performing models are CLIPER, BAM etc. Please refer to http://www.nhc.noaa.gov/modelsummary.shtml for a list of the best performing models that are commonly used.

In this thesis, we present a data mining perspective on m step ahead trajectory prediction problem, where m is for user to choose, but typically 4, 8 and so on. We restrict our attention to hurricanes that lasts at least three days (12 data points), but users may have their choice in the software that we have developed for this chapter. We feel it is a reasonable assumption because ninety-one percent of the hurricanes from 1951 to 2008 last at least three days, so the short-lived hurricanes (less than three days) are rare. The m step ahead trajectory prediction problem has two variants: m step ahead sequential prediction and m step ahead direct prediction. For m step ahead sequential prediction, we predict locations for the upcoming m points one at a time given the initial k points we have observed already. As new data comes in, we constantly adjust the algorithm for the best prediction of the next location. For m step ahead direct prediction, we predict the upcoming m points at one time, given the initial observed k points. k could be arbitrary.

It is obvious that m step ahead direct prediction is a lot harder than m step ahead sequential prediction. Our program can accommodate any k and m, as long as they are in reasonable range. To give an introduction of our algorithms, we first transform the raw longitude and latitude to the Cartesian coordinates for modeling, as the Cartesian coordinates enjoy several advantages over the original scale. Then we implement seven individual algorithms using different data mining techniques including collaborative filtering and time series models or a mixture of both. After collecting the predictions from each individual algorithms, we propose fifteen different model combination and weighting procedures using either static weights or dynamic weights. From the two hurricane examples we will illustrate in Chapter 5, the dynamic updating scheme consistently performs among the best, no matter which individual algorithm outperforms. It has a self selection property. Interestingly but not surprisingly, methods such as collaborative filtering have enjoyed significant improvement as hurricanes become more and more recent, due to the fact that more historical data is being considered in the algorithm. In addition, the prediction error tends to decrease as we predict more recent hurricanes than the past.

At the end of Chapter 5, we present some thoughts on future research directions. One area that seems promising is to modify the collaborating filtering algorithm to leverage the information of more quality past hurricanes, as some past hurricanes may be useful for prediction if we look at the pattern of its entire life cycle rather than the initial few points. Parametric modeling is also an area that may require further attention as hurricanes tend to follow the curvature which starts from the southeast, then moves west, and finally heads towards northeast. Quadratic curve modeling has the potential to improve our algorithm. Lastly, climate covariates may be incorporated in the whole prediction paradigm, as hurricane movement is supposed to closely relate to the local and global climate environment.

Chapter 2

Change Detection In Exponential Family and GLM

This chapter is devoted to a comprehensive treatment of the change detection problem in the exponential family. Section 2.1 contains a brief literature review. Section 2.2 deals with EF-CUSUM statistic derivation. Multivariate Gaussian-CUSUM is discussed as well, with covariance matrix either singular or positive definite. A few examples are given as to how to derive the CUSUM statistic, and Table 2.1 and 2.2 are provided for the convenience of readers. Section 2.2.2 talks about change detection procedure in the generalized linear model context. Section 3.5 contains simulation studies on performance comparison between the EF-CUSUM and the Guassian-CUSUM procedures. The data analysis for Atlantic tropical storms is provided in Section 2.5.

2.1 Literature Review

In this section we provide a partial list of techniques for change detection. As mentioned earlier, some of these originated in industrial quality context, and related methods include Shewhart control charts [129], EWMA control charts [123] and CUSUM [112].

In the context of the CUSUM statistic, which originated from [112] and [113], [91] proposed an asymptotic optimality using the minimax criterion. Later, [107] also established that under Lorden's criterion, when the data is independently and identically distributed with known distributions before and after the change, the CUSUM procedure was indeed optimal. Additionally, [122] showed that CUSUM was Bayesian optimal under Lorden's measure, and [117] derived asymptotic expression for average run length.

The CUSUM technique has been extended to better suit practical needs, including [132] on adaptive CUSUM, [57] on robust average run length with Winsorization, and [90] on transformation of exponential data into approximately normal distribution and compared transformed CUSUM with existing CUSUM procedures. Also, paper [150] proposed transforming serially correlated observations (such as ARMA) into independent, identically distributed sequences while keeping average run length roughly the same. In other directions, paper [96] compared the average run length properties of EWMA with CUSUM, and [6] proved that the CUSUM scheme that utilized BCUSUM mask was uniformly most powerful and compared it with other existing CUSUM procedures. Paper [98] developed robust CUSUM by modifying the likelihood function, [3] proposed CUMIN charts for grouped data and compared CUMIN with CUSUM and Shewhart charts, [16] proposed CUSUM control charts with control limits estimated using bootstrapping when the distribution was unknown, [136] used simultaneous CUSUM control charts to monitor correlated bivariate outcomes in the field of medical research, [23] proposed vector CUSUM and Hotelling T^2 based CUSUM when dealing with multivariate case and compared them to Shewhart scheme, [94] proposed Shewhart-CUSUM scheme to draw advantages of both methods for quick detection of mean change in the normal distribution setting, and [106] extended the approach to binomial data.

Some researchers have treated special cases in the EF-CUSUM family, including [59] on detecting known location and shape change in inverse gamma distribution, [60] on change point detection in unknown mean and variance for normal distribution, [125] used negative binomial CUSUM to study outbreaks of Ross River virus disease and compared it to Early Aberration Reporting System (EARS) CUSUM algorithms, [146] studied large shifts in fraction non-conforming in Poisson CUSUM chart, [93] improved the Poisson CUSUM with FIR and introduced two-in-a-row rule to robust CUSUM. [77] showed optimality of CUSUM for exponential distribution by calculating ARL using Wald's approximation. [61] discussed shift in mean and covariance for multivariate normal distribution using CUSUM, [5] proposed transformation to normality to deal with EF-CUSUM chart, [128] discussed using Kalman Filter and CUSUM to detect residual mean and variance in the regression model, and [119] used rank-based CUSUM procedure to deal with multivariate measurements without normality assumption.

In the context of generalized linear model, papers such as [12] and [73] discussed general linear model, and [86], [20], [118] focused on detecting linear model with different types of error terms.

2.2 Distributional Stability in Exponential Families

2.2.1 Known Parameter Case

Let the data be the random sample $\{X_1, \ldots, X_n\}$, where we know X_1 is observed first, then X_2 is observed, and so on. We assume that $X_1, \ldots, X_\tau$ are identically and independently distributed following an exponential family distribution with probability density or mass function given by

$$p(x; \theta, \phi) = \exp\left\{a(\phi)^{-1}\left(x\theta - b(\theta)\right) + c(x, \phi)\right\}.$$

Here the parameters are θ, which is of the same dimensionality as each of the data-points, and ϕ. We also assume that $X_{\tau+1}, \ldots$ are identically and independently distributed from another exponential family distribution, with probability density function given by

$$\begin{aligned} p(x; \theta + \delta_1, \phi + \delta_2) = \exp\{a(\phi + \delta_2)^{-1}(x(\theta + \delta_1) \\ -b(\theta + \delta_1)) + c(x, \phi + \delta_2)\}. \end{aligned}$$

Here τ is a fixed but unknown parameter denoting the time of change from one distribution to another, and $0 < \tau < \infty$. In the *testing for distributional stability* (TDS) framework we adopt, our interest is in testing the null hypothesis $H_0 : \tau \geq n$ against the alternative hypothesis $H_1 : \tau < n$. We consider all parameter values, other than τ as known constants.

In the paradigm above, we assume τ is unknown while other parameters are known. This stems from the two observations below. First, they made the mathematical formulation simpler and reduced the technical details considerably, and the traditional application domain of statistical process control where CUSUM and related techniques originated treat such parameters as known constants. Additionally, under standard conditions the rate of convergence for the estimated change point to the true change point (if there is one) is faster than those of parameter estimates, consequently the asymptotics of parameter estimators can be fully de-linked from the asymptotics of the stability detection hypothesis test. Nevertheless, assuming some, or all, of these parameters as unknown is an easy extension but requires additional technical conditions and assumptions, which we will discuss in a separate section. Note that the time-ordering of the observations is not an integral part to our methodology. Also, multiple change-points may be allowed. For the former, we would assume that there is some permutation of the data, say $X_{\sigma_1}, \ldots, X_{\sigma_n}$ such that $X_{\sigma_1}, \ldots, X_{\sigma_\tau}$ are independent and identically distributed with some exponential family distribution with parameters θ and

ϕ, while $X_{\sigma_{\tau+1}}, \ldots$ independent and identically distributed with the same distribution with a different set of parameter values. Also, multiple change-points $\tau_1, \ldots, \tau_k$ can be easily accommodated in the above framework, and both the null and alternative hypothesis are made more complex. In other words, we can extend our study to the case where, for some permutation of the indices, the data may be partitioned into k_0 segments under the null and k_1 segments under the alternative. Here, each segment of data is a set of independent, identically distributed exponential family random variables with its own distinct set of parameters. Our current problem may be thought of as the special case where $\sigma_i = i$ for $i = 1, \ldots, n$, $k_0 = 1$ and $k_1 = 2$. Extensions like those described above may lead to new approaches for solving several problems in applied statistics. However, in the interest of clarity of presentation, and to keep this thesis at a reasonable length, we do not pursue such extensions here. Another important extension from our methodology is the temporal dependence structure of the data. Our method has a natural extension to time series and other dependent data with potential change points, for which the likelihood can be written and computed, and an equivalent CUSUM testing framework can be established.

In our first result below, we obtain the test statistic for the hypothesis test described above. We adopt the convention that $\sum_{i=a}^{b} Y_i = 0$ whenever $a > b$, for any sequence of (possibly random) reals $\{Y_i\}$.

Theorem 2.2.1. *Let*

$$\begin{aligned} Y_i &= a(\phi+\delta_2)^{-1}\left(X_i(\theta+\delta_1) - b(\theta+\delta_1)\right) + c(X_i, \phi+\delta_2) \\ &\quad - a(\phi)^{-1}\left(X_i\theta - b(\theta)\right) - c(X_i, \phi), \end{aligned}$$

for $i = 1, \ldots, n$, and further define $S_k = \sum_{i=1}^{k} Y_i$, adopting the convention that $S_0 = 0$.

The likelihood ratio test statistic for testing the null hypothesis $H_0 : \tau \geq n$ against the alternative hypothesis $H_1 : \tau < n$ is given by $T_n = S_n - \min_{0 \leq k < n} S_k$, and the null hypothesis is rejected if $T_n \geq L$ for some constant L.

Proof. The likelihood function under the null is:

$$L_0(\tau) = \prod_{i=1}^{n} \exp\left\{a(\phi)^{-1}(x_i\theta - b(\theta)) + c(x_i, \phi)\right\}.$$

The likelihood function under the alternative is:

$$\begin{aligned} L_1(\tau) &= \prod_{i=1}^{\tau} \exp\left\{a(\phi)^{-1}(x_i\theta - b(\theta)) + c(x_i,\phi)\right\} \\ &\times \prod_{i=\tau+1}^{n} \exp\left\{a(\phi+\delta_2)^{-1}(x_i(\theta+\delta_1) - b(\theta+\delta_1)) + c(x_i,\phi+\delta_2)\right\}. \end{aligned}$$

For a likelihood ratio test, we use

$$\begin{aligned} \Lambda &= \frac{\max_{0\le\tau<n} L_1(\tau)}{\max_{\tau\ge n} L_0(\tau)} \\ &= \max_{\tau\ge n} \prod_{i=1}^{n} \exp\left\{a(\phi)^{-1}(x_i\theta - b(\theta)) + c(x_i,\phi)\right\}^{-1} \\ &\times \max_{0\le\tau<n} \prod_{i=1}^{\tau} \exp\left\{a(\phi)^{-1}(x_i\theta - b(\theta)) + c(x_i,\phi)\right\} \\ &\times \prod_{i=\tau+1}^{n} \exp\left\{a(\phi+\delta_2)^{-1}(x_i(\theta+\delta_1) - b(\theta+\delta_1)) + c(x_i,\phi+\delta_2)\right\} \\ &\propto \max_{0\le\tau<n} \prod_{i=1}^{\tau} \exp\left\{a(\phi)^{-1}(x_i\theta - b(\theta)) + c(x_i,\phi)\right\} \\ &\times \prod_{i=\tau+1}^{n} \exp\left\{a(\phi+\delta_2)^{-1}(x_i(\theta+\delta_1) - b(\theta+\delta_1)) + c(x_i,\phi+\delta_2)\right\} \end{aligned}$$

To maximize Λ, it is equivalent to maximizing

$$\begin{aligned} & \sum_{i=1}^{\tau} \left\{a(\phi)^{-1}(x_i\theta - b(\theta)) + c(x_i,\phi)\right\} \\ + & \sum_{i=\tau+1}^{n} \left\{a(\phi+\delta_2)^{-1}(x_i(\theta+\delta_1) - b(\theta+\delta_1)) + c(x_i,\phi+\delta_2)\right\} \\ = & \sum_{i=1}^{\tau} \left\{a(\phi)^{-1}(x_i\theta - b(\theta)) + c(x_i,\phi)\right\} \\ + & \sum_{i=1}^{n} \left\{a(\phi+\delta_2)^{-1}(x_i(\theta+\delta_1) - b(\theta+\delta_1)) + c(x_i,\phi+\delta_2)\right\} \\ - & \sum_{i=1}^{\tau} \left\{a(\phi+\delta_2)^{-1}(x_i(\theta+\delta_1) - b(\theta+\delta_1)) + c(x_i,\phi+\delta_2)\right\} \\ = & \sum_{i=1}^{n} \left\{a(\phi+\delta_2)^{-1}(x_i(\theta+\delta_1) - b(\theta+\delta_1)) + c(x_i,\phi+\delta_2)\right\} \\ - & \sum_{i=1}^{\tau} \left\{a(\phi+\delta_2)^{-1}(x_i(\theta+\delta_1) - b(\theta+\delta_1)) + c(x_i,\phi+\delta_2) - a(\phi)^{-1}(x_i\theta - b(\theta)) - c(x_i,\phi)\right\} \end{aligned}$$

Define $y_i = a(\phi+\delta_2)^{-1}(x_i(\theta+\delta_1)-b(\theta+\delta_1))+c(x_i,\phi+\delta_2)-a(\phi)^{-1}(x_i\theta-b(\theta))-c(x_i,\phi)$, and $S_k = \sum_{i=1}^{k} y_i$. we choose $\hat{\tau} = \arg\min_{0\le k<n} S_k$.

Therefore the statistic Λ is given by

$$\begin{aligned}
\Lambda(\hat{\tau}) &= \frac{L_1(\hat{\tau})}{L_0(\hat{\tau})} \\
&= \prod_{i=1}^{\hat{\tau}} \exp\{a(\phi)^{-1}(x_i\theta - b(\theta)) + c(x_i,\phi)\}^{-1} \prod_{i=\hat{\tau}+1}^{n} \exp\{a(\phi)^{-1}(x_i\theta - b(\theta)) + c(x_i,\phi)\}^{-1} \\
&\times \prod_{i=1}^{\hat{\tau}} \exp\left\{a(\phi)^{-1}(x_i\theta - b(\theta)) + c(x_i,\phi)\right\} \\
&\times \prod_{i=\hat{\tau}+1}^{n} \exp\left\{a(\phi+\delta_2)^{-1}(x_i(\theta+\delta_1) - b(\theta+\delta_1)) + c(x_i,\phi+\delta_2)\right\} \\
&= \prod_{i=\hat{\tau}+1}^{n} \left(\exp\{a(\phi)^{-1}(x_i\theta - b(\theta)) + c(x_i,\phi)\}\right)^{-1} \\
&\times \prod_{i=\hat{\tau}+1}^{n} \exp\left\{a(\phi+\delta_2)^{-1}(x_i(\theta+\delta_1) - b(\theta+\delta_1)) + c(x_i,\phi+\delta_2)\right\} \\
&= \exp\{\sum_{i=\hat{\tau}+1}^{n} y_i\} \\
&= \exp\{S_n - \min_{0\le k<n} S_k\}
\end{aligned}$$

We reject if $S_n - \min_{0\le k<n} S_k \ge L$ for some L. □

Note that the test statistic T_n may be written recursively as $T_n = \max\{0, T_{n-1}+Y_n\}$, with $T_0 = 0$. This form is reminiscent of the the celebrated CUSUM statistic. In view of this, we call T_n the *exponential family CUSUM* statistic. We obtain the classical CUSUM statistic as a special case in Corollary 2.2.1 below.

Remark 2.2.1

Recall that we reject the null hypothesis $H_0 : \tau \ge n$ if $T_n \ge L$ for some L. The standard method for choosing L in the hypothesis testing paradigm is by controlling the probability of Type-I error at some pre-determined level α. However, in the sequential statistics literature, the comparable technique is to control the expectation of the *run length* under the null hypothesis. The run length R is defined as the number of observations gathered before a decision is reached on the rejection or acceptance of the null hypothesis. In the present framework, we have $R = \inf\{n : S_n - \min_{0\le k<n} S_k = T_n \ge L\}$. The value of L may be obtained by fixing the value of $\mathbb{E}R (= ARL)$ assuming $\tau = \infty$, at a pre-determined value ARL_0. The notation ARL stands for *average run length*. For ease in comparison with

existing procedures for change detection, we will report expected run length $\mathbb{E}\max(R-\tau, 0)$ under the alternative as a measure of power, the way it is done in sequential statistics literature. The probability of Type-I (Type-II) error and the expected run length under the null (alternative) hypothesis are related, though the relation is generally not easy to obtain.

To sum up, L is the standard critical value that is used to compare the test statistic with in a hypothesis testing problem, and ARL_0 is related to the probability of Type-I error. Thus, these quantities are versions of quantities that arise in the standard hypothesis testing protocol. We use L and ARL_0 in place of a standard critical value and level of a test because of the historic relation of the problem being addressed in this thesis (hypothesis testing for stability) to the literature on process monitoring and change detection.

The choice of L determines the ARL_0. In general, the larger the L, the larger ARL_0, thus the smaller Type-I error, and the larger Type-II error. We choose ARL_0 pretty much in the same philosophy as choosing the level of the test α. In the process monitoring and change detection paradigm, $ARL_0 = 1/\alpha$.

Note that $T_n \geq 0$ almost surely, hence a non-trivial test is obtained only when L is strictly positive. Our next result shows that this relation is fairly easy to ensure in practice.

Theorem 2.2.2. $\mathbb{E}_{\tau=\infty}R(= ARL_0) \geq 1$ *if and only if the critical value L is positive.*

Proof of Theorem 2.2.2 *The necessity part:* If $L \leq 0$, since $R = \inf\{n : S_n - \min_{0\leq k<n} S_k \geq L\}$, we have $S_0 - \min_{0\leq k<0} S_k = 0 \geq L$. Hence we have $R = 0$ almost surely, and therefore $\mathbb{E}_{\tau=\infty}(R) = 0$, which is contradictory to $ARL_0 \geq 1$. *The sufficiency part:* If $L > 0$, then R cannot be zero because $S_0 - \min_{0\leq k<0} S_k = 0 < L$, hence R is at least 1. Therefore $ARL_0 \geq 1$. □

We now state some special cases of Theorem 2.2.1, which are of interest. Our first such result deals with the case where the observations are Normally distributed. We use the notation $\overset{\text{iid}}{\sim}$ for independent and identically distributed.

Corollary 2.2.1. Suppose $X_1, \ldots, X_\tau \overset{\text{iid}}{\sim} N(\mu, \sigma_1^2)$ and $X_{\tau+1}, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu + \delta_1, \sigma_2^2)$. For testing the null hypothesis $H_0 : \tau \geq n$ against the alternative $H_1 : 0 \leq \tau < n$, the likelihood ratio statistic is given by $C_n = S_n - \min_{0\leq k<n} S_k$, where $S_k = \sum_{i=1}^{k} Y_i$ and

$$\begin{aligned} Y_i &= \log(\sigma_1) + \frac{1}{2}\sigma_1^{-2}(X_i - \mu)^2 \\ &\quad - \log(\sigma_2) - \frac{1}{2}\sigma_2^{-2}(X_i - \mu - \delta_1)^2. \end{aligned}$$

We omit the proof of this Corollary, which follows easily from Theorem 2.2.1. In the very special case where $\sigma_1 = \sigma_2 = 1, \mu = 0$, we obtain $Y_i = (X_i - \delta/2)$, and hence obtain $S_n - \min_{0 \le k < n} S_k = C_n = \max\{0, C_{n-1} + X_i - \delta/2\}$, with $C_0 = 0$. This expression is that of the classical Gaussian-CUSUM, where the factor $\delta/2$ is often called the *allowance constant*.

The statistic C_n defined as $C_n = \max\{0, C_{n-1} + X_i - \delta/2\}$ (with $C_0 = 0$) is often used as a default statistic for change detection. Our result above shows that this statistic may also be obtained in a non-sequential framework, however, the assumption of normal distribution seems unavoidable. Since C_n is used for change detection in non-normal data also, it is of interest to know under what circumstance it may obtain reasonable accuracy and precision with change detection. Our next theorem describes the conditions under which using C_n as a statistic may be a reasonable procedure.

Theorem 2.2.3. *Consider the framework of Theorem 2.2.1. In addition, assume that the third derivative of $b(\cdot)$ at θ_0 is zero, i.e., $b'''(\theta_0) = 0$, that δ_1 is small and $\delta_2 = 0$. Under these assumptions, the difference between the Normality-based CUSUM C_n and the exponential family CUSUM T_n is as follows: $|C_n - T_n| = o(n\delta_1)$.*

Remark 2.2.2
In the above theorem, the symbol o means for any n, $|C_n - T_n|/(n\delta_1) \to 0$ as $\delta_1 \to 0$ in the mathematical limiting sense.

In the case of binomial distribution with parameter p, the natural parameter is $\theta = \log((1-p)^{-1}p)$, and $b(\theta) = n\log(1+\exp(\theta))$, ϕ is taken as a constant. Also $b'''(\theta) = (1+\exp(\theta))^{-4}\{n\exp(\theta)(1+\exp(\theta))(1-\exp(\theta))\}$, $b'''(\theta_0) = 0$ iff $\theta_0 = 0$. In that case, $p = \frac{1}{2}$. To conclude, when $p = \frac{1}{2}$, a change from $p \to p + \delta_1$ using Gaussian-CUSUM $\tilde{y}$ and exponential family CUSUM y yield similar performance in the sense that $|\tilde{y} - y| = o(\delta_1)$.

Corollary 2.2.2. For the same detection problem as above, under the condition of $b'''(\theta_0) = b''''(\theta_0) = 0$, δ_1 is small and $\delta_2 = 0$, we get an even stronger result $|\tilde{y} - y| = o(\delta_1^2)$.

Example 2.2.1.1
Change from $N_p(\mu, \Sigma_1)$ to $N_p(\mu + \delta, \Sigma_2)$

Multivariate normal distribution is commonly used in multi-dimensional setting. The CUSUM for multivariate normal distribution is somewhat more complicated, and therefore we divide this problem into the following cases based on the nature of the variance-covariance matrix. In all the cases listed below, the test statistic is $C_n = S_n - \min_{0 \le k < n} S_k$, where $S_k = \sum_{i=1}^{k} Y_i$ and Y_i depends from one case to another.

1. $\Sigma_1 = \Sigma_2 = \Sigma$, *where* Σ *is positive definite*. Based on the following density function: $f(x|\mu, \Sigma) = (2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\}$ it is straightforward to derive

the CUSUM statistic based on $Y_i = (x_i - \mu - \frac{1}{2}\delta)'\Sigma^{-1}\delta$. If we let $p = 1$, we are back to the univariate normal situation.

2. $\Sigma_1 = \Sigma_2 = \Sigma$, *where* Σ *is a singular.*

   Assume rank $(\Sigma) = r, r < p$. By linear algebra, there exists an orthogonal matrix Q_{p*p}, such that $Q\Sigma Q' = \Lambda$, where Λ is $\begin{pmatrix} \lambda_1 & & & & & & \\ & \cdot & & & & & \\ & & \cdot & & & & \\ & & & \lambda_r & & & \\ & & & & 0 & & \\ & & & & & \cdot & \\ & & & & & & \cdot \\ & & & & & & & 0 \end{pmatrix}_{p\times p}$. Here $\lambda_i > 0, i = 1, 2...r$. So $Z = QX \sim N_p(Q\mu, \Lambda)$. Let P be the matrix $\begin{pmatrix} 1 & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & 1 & \end{pmatrix}_{r\times p}$.

   Let $K = PZ \sim N_r(PQ\mu, \tilde{\Lambda})$, where $\tilde{\Lambda}$ is $\begin{pmatrix} \lambda_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \lambda_r \end{pmatrix}_{r\times r}$. So the problem is reduced to a change of $N_r(PQ\mu, \tilde{\Sigma})$ to $N_r(PQ(\mu+\delta), \tilde{\Sigma})$, and we are back to case 1. The CUSUM statistic is based on $Y_i = (x_i - \mu - \frac{1}{2}\delta)'(PQ)'\tilde{\Sigma}^{-1}PQ\delta$.

3. $\Sigma_1 \neq \Sigma_2$, *where* Σ_1, Σ_2 *are both positive definite.* Following previous discussion, the CUSUM statistic is based on $Y_i = \frac{1}{2}\log(|\Sigma_1|^{-1}|\Sigma_2|) + \frac{1}{2}(x_i - \mu - \delta)'\Sigma_2^{-1}(x_i - \mu - \delta) - \frac{1}{2}(x_i - \mu)'\Sigma_1^{-1}(x_i - \mu)$.

4. $\Sigma_1 \neq \Sigma_2$, *where* Σ_1, Σ_2 *are both singular.*

   Based on discussion of case 2, our CUSUM statistic is based on $Y_i = (\frac{r_2}{2} - \frac{r_1}{2})\log(2\pi) + \frac{1}{2}\log(|\tilde{\Lambda_1}|^{-1}|\tilde{\Lambda_2}|) - \frac{1}{2}(P_1Q_1(x_i-\mu))'\tilde{\Lambda_1}^{-1}(P_1Q_1(x_i-\mu)) + \frac{1}{2}(P_2Q_2(x_i-\mu-\delta))'\tilde{\Lambda_2}^{-1}(P_2Q_2(x_i - \mu - \delta))$. Here P_1, Q_1, P_2, Q_2 are such that $P_1Q_1\Sigma_1Q_1'P_1' = \tilde{\Lambda_1}$, $P_2Q_2\Sigma_2Q_2'P_2' = \tilde{\Lambda_2}$, and rank ($\tilde{\Lambda_1}$) = rank (Σ_1), rank($\tilde{\Lambda_2}$) = rank (Σ_2), $\tilde{\Lambda_1}, \tilde{\Lambda_2}$ are $r_1 \times r_1$ and $r_2 \times r_2$ diagonal matrix.

5. $\Sigma_1 \neq \Sigma_2$, *where* Σ_1 *is positive definite,* Σ_2 *is singular.* In this case we have $Y_i = \frac{r_2 - p}{2}\log(2\pi) + \frac{1}{2}\log(|\tilde{\Lambda_1}|^{-1}|\tilde{\Lambda_2}|) + \frac{1}{2}(P_2Q_2(x_i - \mu - \delta))'\tilde{\Lambda_2}^{-1}(P_2Q_2(x_i - \mu - \delta))$
   $- \frac{1}{2}(x_i - \mu)\Sigma_1^{-1}(x_i - \mu)$, where $P_2Q_2\Sigma_2Q_2'P_2' = \tilde{\Lambda_2}$, rank($\tilde{\Lambda_2}$) = rank (Σ_2), $\tilde{\Lambda_2}$ is $r_2 \times r_2$ diagonal matrix.

2.2.2 Generalized Linear Model and CUSUM

In this section, we consider data of the form $(y_1, \mathbf{x_1}), \ldots, (y_n, \mathbf{x_n})$. Here, the y_i's are the responses, and the $\mathbf{x}_i$'s are covariates that are considered to be fixed constant vectors. We assume that $y_i's$ come from the distribution $p(y_i|\theta_i) = \exp\{a(\phi)^{-1}(y_i\theta_i - b(\theta_i)) + c(y_i, \phi)\}$, where $\theta_i = \mathbf{x_i}'\beta$ is the canonical parameter under stable distributional regime and $a(\phi) > 0$ is a dispersion parameter. Our main result below generalizes the main result of the previous section, and presents change detection test statistic for generalized linear models.

Theorem 2.2.4. *Assume that $(y_1, \mathbf{x_1}), \ldots, (y_\tau, \mathbf{x}_\tau)$, the true model is $\theta_i = \mathbf{x_i}'\beta$, and for $(y_{\tau+1}, \mathbf{x_{\tau+1}}), \ldots, (y_n, \mathbf{x_n})$, the true model is $\theta_i = \mathbf{x_i}'(\beta + \delta)$, where β, δ is known. For the hypothesis testing $H_0 : \tau \geq n$ vs $H_1 : 0 \leq \tau < n$. If we denote $z_i = y_i\mathbf{x_i}'\delta - b(\mathbf{x_i}'(\beta + \delta)) + b(\mathbf{x_i}'\beta)$ and $S_k = \sum_{i=1}^{k} z_i$, then the rejection region is $S_n - \min_{0 \leq k < n} S_k \geq L$.*

2.2.3 Estimated Parameter Case

We now illustrate the results presented above extend to the case where the parameters are unknown. For simplicity of presentation, we omit the scaling function $a(\phi)$ for the first two results below. We begin with the single parameter framework where $X_1, \ldots, X_{\tau_n}$ are independent and identically distributed with density

$$p(x; \theta_0) = \exp\left\{(x\theta_0 - b(\theta_0)) + c(x)\right\},$$

and $X_{\tau_n+1}, \ldots$ are i.i.d. with density

$$p(x; \theta_1) = \exp\left\{(x\theta_1 - b(\theta_1)) + c(x)\right\}.$$

We assume $\theta_1 \neq \theta_0$ throughout. We test the null hypothesis $H_0 : \tau_n \geq n$ against the alternative $H_1 : 0 \leq \tau_n < n$. Let us denote the maximum likelihood estimator for θ_0 based on $X_1, \ldots, X_n$ as $\hat{\theta}_{00}$; note that this is under the null hypothesis scenario. Also, under the alternative hypothesis scenario, the likelihood $L(\theta_0, \theta_1, \tau_n) = \prod_{i=1}^{\tau_n} p(X_i; \theta_0) \prod_{i=\tau_n+1}^{n} p(X_i; \theta_0)$ is maximized at $(\hat{\theta}_{10}, \hat{\theta}_{11}, \hat{\tau}_n)$. We have the following result:

Theorem 2.2.1. *In the framework described above, the likelihood ratio test statistic is given by*

$$\begin{aligned} T_{n1} \quad = \quad & (\hat{\theta}_{10} - \hat{\theta}_{00}) \sum_{i=1}^{\hat{\tau}_n} X_i + (\hat{\theta}_{11} - \hat{\theta}_{00}) \sum_{i=\hat{\tau}_n+1}^{n} X_i \\ & -\hat{\tau}_n b(\hat{\theta}_{10}) - (n - \hat{\tau}_n) b(\hat{\theta}_{11}) + n b(\hat{\theta}_{00}). \end{aligned}$$

Further, under either $\tau_n \geq n$ or $\tau_n/n \in (0,1)$, the parametric bootstrap scheme may be used to estimate the distribution of T_{n1}, and consequently obtain a rejection region and p-value of the above hypothesis test.

It may be noted, however, that the above test statistic can suffer from extremely low power, depending on the values of θ_0, θ_1 and τ_n. One reason for this performance deficiency is that θ_{00} is not a consistent estimator for θ_0 under the alternative hypothesis. In order to address this issue and improve the performance capabilities of our testing procedure, we propose a modification of the usual likelihood ratio test, whereby we use $\hat{\theta}_{10}$ as the estimator for θ_0, even under the null hypothesis. We have the following result:

Theorem 2.2.2. *In the framework of Theorem 2.2.1, the profile likelihood ratio test statistic is*

$$T_{n2} = (\hat{\theta}_{11} - \hat{\theta}_{00}) \sum_{i=\hat{\tau}_n+1}^{n} X_i - (n - \hat{\tau}_n)(b(\hat{\theta}_{11}) - b(\hat{\theta}_{00})).$$

Further, under either $\tau_n \geq n$ or $\tau_n/n \in (0,1)$, the parametric bootstrap scheme may be used to estimate the distribution of T_{n1}, and consequently obtain a rejection region and p-value of the above hypothesis test. Further, the power of this test tends to one when $\tau_n/n \in (0,1)$. In addition, $(\hat{\theta}_{10}, \hat{\theta}_{11}, \hat{\tau}_n)$ converge in probability to $(\theta_0, \theta_1, \tau_n)$ under standard conditions.

The above test statistic can be obtained from the profile likelihood (for null and alternative), when θ_0 is replaced with $\hat{\theta}_{10}$. Another useful variant is the case where both θ_0 and θ_1 may be estimated from the full data, perhaps under some restrictions on the model. An example is where the the null distribution is $N(\theta_0, \sigma^2)$, and after τ_n it changes to $N(\theta_0 + c\sigma, \sigma^2)$ for some known constant c. This formulation is particularly useful for applications, where it may be of importance to detect only practically significant lack of stability of distributions, and not just statistically significant ones. In our simulation examples and the real data analysis below, we consider the above specification where we test for a change in mean in terms of c standard deviation units. We study results with $c = 1, 1/2, 1/4$ as potential cases of relatively easy, not easy and hard change-detection scenarios. This framework is adopted since it makes sense to describe the distance between the null and alternative scenarios in terms of "units of standard deviation". Also, in samples of finite sizes, the only scenario where we get reasonable power in hypothesis tests is when the two hypotheses are sufficiently apart. Additionally, for practical purposes, even if there is a change but the change is minute and negligible, the hypotheses test may be redundant. Based on all these considerations, it is advisable to test hypotheses that are a reasonable number of standard deviation units away from each other.

There can be several other results relating to stability detection with estimated parameters, under various assumptions and technical conditions, which we will address in future work. We conclude this section with a result on stability detection when parameters are estimated in a generalized linear model.

Theorem 2.2.3. *Assume that* $(y_1, \mathbf{x_1}), \ldots, (y_{\tau_n}, \mathbf{x}_{\tau_n})$, *the true model is* $\theta_{i0} = \mathbf{x_i}'\beta_0$, *and for* $(y_{\tau_n+1}, \mathbf{x_{\tau_n+1}}), \ldots, (y_n, \mathbf{x_n})$, *the true model is* $\theta_{i1} = \mathbf{x_i}'\beta_1$. *For the hypothesis testing* $H_0 : \tau_n \geq n$ *vs* $H_1 : 0 \leq \tau_n < n$, *the test statistic is*
$T_{n3} = \sum_{i=1\hat{\tau}_n+1}^{n} a^{-1}(\hat{\phi}) \left\{ y_i \mathbf{x_i}' (\hat{\beta}_1 - \hat{\beta}_0) - b(\mathbf{x_i}'\beta_1) + b(\mathbf{x_i}'\beta_0) \right\}$

We present below a sketch of the proof of the above result.

Sketch of Proof of Theorem 2.2.3

The likelihood function under the alternative hypothesis is

$$\begin{aligned}
& L_1(\beta_0, \beta_1, \tau_n, \phi) \\
= & \prod_{i=1}^{\tau_n} \exp\{a(\phi)^{-1}(y_i \mathbf{x_i}'\beta_0 - b(\mathbf{x_i}'\beta_0)) + c(y_i, \phi)\} \\
& \times \prod_{i=\tau_n+1}^{n} \exp\{a(\phi)^{-1}(y_i \mathbf{x_i}'\beta_1 - b(\mathbf{x_i}'\beta_1)) + c(y_i, \phi)\}.
\end{aligned}$$

Suppose this function is maximized at $(\hat{\beta}_0, \hat{\beta}_1, \hat{\tau}_n, \hat{\phi})$. We evaluate the likelihood under the null hypothesis at $\hat{\beta}_0, \hat{\phi}$, and obtain the profile likelihood ratio as

$$\begin{aligned}
\Lambda(\tau) &= \frac{L_1(\hat{\beta}_0, \hat{\beta}_1, \hat{\tau}_n, \hat{\phi})}{L_0(\hat{\beta}_0, \hat{\phi})} \\
&= \exp\left[\sum_{i=1\hat{\tau}_n+1}^{n} a^{-1}(\hat{\phi}) \left\{ y_i \mathbf{x_i}' (\hat{\beta}_1 - \hat{\beta}_0) \right.\right. \\
&\qquad \left.\left. - b(\mathbf{x_i}'\beta_1) + b(\mathbf{x_i}'\beta_0) \right\} \right].
\end{aligned}$$

$\square$

In the generalized linear model case also, the parametric bootstrap is a viable way of approximating the distribution of T_{n3}, and thus eliciting the properties of the test for stability.

2.3 Simulation Study

In this section, we discuss a simulation study on the change of parameter(s) for binomial, exponential, gamma and poisson distributions, and compare the EF-CUSUM statistic with the Gaussian-CUSUM statistic, under the constraint that the mean and the standard deviation of both distributions are equal. Based on the exponential family density $f(x;\theta,\phi) = \exp\{a(\phi)^{-1}(x\theta - b(\theta)) + c(x,\phi)\}$, it is easy to calculate $E(X) = b'(\theta)$, and $var(X) = b''(\theta)a(\phi)$. When there is change in parameter from θ to $\theta + \delta_1$ and from ϕ to $\phi + \delta_2$, we have $E(X) = b'(\theta + \delta_1)$ and $var(X) = b''(\theta + \delta_1)a(\phi + \delta_2)$. So the corresponding Gaussian assumption-based setting is a change from $N(b'(\theta), b''(\theta)a(\phi))$ to $N(b'(\theta + \delta_1), b''(\theta + \delta_1)a(\phi + \delta_2))$.

The simulation procedure can be described as follows: First, we control false alarms by carefully choosing L under the null distribution. We generate $\{x_n\}_{n=1}^{T=2000} \overset{\text{iid}}{\sim} f(x|\theta)$ for 2500 times. Here T is fixed at 2000 for illustration. The density $f(x|\theta)$ is a distribution belonging to the appropriate exponential family. Define $R = \inf\{i : S_i - \min_{0 \le k < i} S_k \ge L\}$, where $S_k = \sum_{i=1}^{k} y_k$ is the EF-CUSUM statistic as we derived. For a fixed L, and for each simulation, we can compute a value of R. Its expectation $E(R)$ can be computed based on these 2500 simulations. Fixing $ARL_0 = 200$, we obtain L such that $\frac{|E_0(R)-200|}{200}$ is minimized. Since $E_0(R)$ is an increasing function of L, with values ranging from 0 to ∞, such an L exists, and is unique.

Second, we compute $E((R-\tau)^+)$ under the alternative distribution. Let τ be an unknown parameter. Again we simulate $x_1, \dots, x_\tau \overset{\text{iid}}{\sim} f(x|\theta)$ and $x_{\tau+1}, \dots, x_T \overset{\text{iid}}{\sim} f(x|\theta + \delta)$ for 2500 times, where δ is known. For each $\tau = 0, 1, \dots, 100$, use the L in the first step and compute R for 2500 times to get the mean, median, standard deviation, and maximum of $(R(\tau))$. Finally, we repeat the same procedure for the normal case, compute $E_1(R(\tau))$ (and other summary statistics) for the Gaussian-CUSUM and compare it with $E_1(R(\tau))$ for the exponential family CUSUM.

From the simulation results in Tables 2.3-2.5 and Figure 2.1, one key finding is that in most cases, the EF-CUSUM statistic performs better than the Gaussian-CUSUM statistic except for one occasion when the underlying distribution is exponential distribution. Here performance is based on the mean of the run length after the change time τ until a signal occurs. Also note that for small shift in parameter, exponential CUSUM has a considerable advantage over the Gaussian-CUSUM, while for large shift in parameter, the EF-CUSUM still works better than the Gaussian-CUSUM, but not significantly different.

We also discover that $E_1(R(\tau))$ does not vary a lot with τ changing from 0 to 100 for a particular distribution in the exponential family. Particularly, for τ close to 0 or close to 100, $E_1(R(\tau))$ is still quite stable. In the table, we showed the $E_1(R(99))$ as a representative of the performance for the statistic. In addition, the median, standard deviation and maximum of average run length tell the same story as the mean.

2.4 When More Parameters of Interest Are Unknown

In the previous sections, we discuss our methodology under the framework that the unknown parameter is the change point τ, and all other parameters are known. Under such, we are able to show that the EF-CUSUM based approach performs better than the Normal-CUSUM approach if the underlying distribution is not normal. Simulation studies compare different scenarios in the exponential family framework, and consistently show the above.

When more parameters are involved, for example, in the parametric model, we have the unknown parameters θ in the model as well as the unknown change point τ. Asymptotic results appear in some research papers, eg, [62], which say the normalized likelihood ratio statistic follows the Gumbel distribution, a type of extreme value distribution. However, the convergence is slow, and may not always provide good approximation.

In this paper, we are going to propose a parametric bootstrap approach, which is a natural extension of the proposed framework. Firstly, as is traditionally treated, we are going to assume the first m observations are i.i.d, therefore they follow the same exponential family distribution with parameters (θ, ϕ).

$$p(x;\theta,\phi) = \exp\left\{a(\phi)^{-1}\left(x\theta - b(\theta)\right) + c(x,\phi)\right\}.$$

The mean is expressed as $b'(\theta)$ and the variance is $b''(\theta)a(\phi)$. The change point τ is assumed to be greater than m. Based on the first m observations, we estimate the parameters using either the maximum likelihood estimation or the method of moments approach. In order for the problem to be interesting, we assume the new distribution (if there is any) follows an exponential family of the same type, but with a different set of parameters (θ', ϕ'). The new set of parameters is considered to be c standard deviations away from the original set of parameters (θ, ϕ). Here c can be positive or negative, and can be adjusted depending on the problem. In the Poisson distribution case, for example, if the mean of the first m observations is $\theta = \lambda$, since the mean and variance is the same for Poisson distribution, we set $\theta' = c\theta^{1/2} = c\lambda^{1/2}$. In the next step, we perform the EF-CUSUM methodology for a change from (θ, ϕ) to (θ', ϕ') according to Section 2.2. This is essentially a parametric boostrap, which is a computationally powerful algorithm when theoretical results are hard to obtain or hard to compute. In this step, just as the simulation study suggests, we generate the null distribution based on estimated parameters (θ, ϕ), and find the cutoff for Type I error. Finally, from the real data, we perform the methodology and draw conclusion.

2.5 Hurricane Analysis with Exponential Family

We now discuss a case study of Atlantic tropical storms, for which data is available for every six hours from its inception till finish. For each storm, the following information is recorded: date and time, hurricane identity, hurricane name, position in latitude and longitude, maximum sustained winds in knots, and central pressure in millibars.

We present our results from three studies on Atlantic hurricanes here. Each of these studies are carried out on two data sets: a longer series from 1851-2008 and a shorter series from 1951-2008. The expectation-maximization algorithm was used for missing data segments in the longer series when required, this problem does not arise in the shorter series.

First, we consider the problem of testing for distributional stability for the yearly number of hurricanes between 1851-2008. This yearly data is modeled as independent Poisson($\hat{\mu}$), and a potential change to Poisson($\hat{\mu} + \delta$) is studied. The choice of Poisson distribution is due to the following two reasons: first, it is count data, and normal assumption does not seem appropriate here. As a matter of fact, we conduct the Shapiro-Wilk test of normality, and the p-value is less than 10^{-6}, while the goodness of fit test for Poisson distribution shows that the p-value is greater than 0.05, which indicates no evidence of rejection of the Poisson distribution. The independence stems from the partial autocorrelation plot, with no evidence of dependence in the first five lags. See Figure 2.4. We assume that any potential change point occurred after 1900, and use the data previous to it for estimating parameters. (We have also considered other segments of time and have performed several checks around the different tuning parameter choices we have made. The results are largely invariant to such choices). We estimate $\hat{\mu} = 7.54$, and fix $\delta = c\hat{\sigma}$, where c is predetermined as $\frac{1}{4}$, $\frac{1}{2}$ and 1, and $\hat{\sigma} = 2.75$ is the estimated standard deviation. Note that $\sigma \approx \mu^{\frac{1}{2}}$ because for the Poisson distribution, the mean equals the variance. Then we create the Poisson CUSUM statistic as given in Table 2.1. We get L based on $E_0(R) = 200$, and search for the first n that satisfies $S_n - \min_{0 \leq k < n} S_k \geq L$ with the hurricane data.

In view of the fact that the data from the 19^{th} century and the first half of the 20^{th} century may not be entirely reliable, we repeated the above analysis on detecting change for the Atlantic tropical storms from year 1951 to 2008. We assume that the potential change could only occur after 1970. For detecting potential change Poisson($\hat{\mu}$) to Poisson($\hat{\mu}+\delta$), we now have $\hat{\mu}=9.8$, and $\delta=c\hat{\sigma}$, where c is predetermined as $\frac{1}{4}$, $\frac{1}{2}$ and 1, and $\hat{\sigma}=2.97$.

The second study has two parts. For the data from 1851-2008, we model the maximum sustained winds and maximum central pressure as $N_2(\hat{\mu},\hat{\Sigma})$, and study potential change to $N_2(\hat{\mu}+\delta,\hat{\Sigma})$. We estimate the mean $\hat{\mu}$ and variance-covariance matrix $\hat{\mathbf{\Sigma}}$ based on the first 50 observations. Here $\hat{\mu}=\begin{pmatrix}104.8\\982.99\end{pmatrix}$, and $\hat{\Sigma}=\begin{pmatrix}\hat{\sigma_{11}} & \hat{\sigma_{12}}\\ \hat{\sigma_{21}} & \hat{\sigma_{22}}\end{pmatrix}=\begin{pmatrix}199.96 & -20.66\\ -20.66 & 367.56\end{pmatrix}$. Let $\delta=\begin{pmatrix}c\hat{\sigma_{11}}\\ c\hat{\sigma_{22}}\end{pmatrix}$, where c is predetermined as $\frac{1}{4}$, $\frac{1}{2}$ and 1.

In a variation of the second study, we consider maximum sustained wind speed and *minimum* central pressure as $N_2(\hat{\mu},\hat{\Sigma})$ and study potential change to $N_2(\hat{\mu}+\delta,\hat{\Sigma})$. Here $\hat{\mu}=\begin{pmatrix}129.5\\937.6\end{pmatrix}$, and $\hat{\Sigma}=\begin{pmatrix}\hat{\sigma_{11}} & \hat{\sigma_{12}}\\ \hat{\sigma_{21}} & \hat{\sigma_{22}}\end{pmatrix}=\begin{pmatrix}376.05 & -220.47\\ -220.47 & 237.41\end{pmatrix}$. Let $\delta=\begin{pmatrix}c\hat{\sigma_{11}}\\ c\hat{\sigma_{22}}\end{pmatrix}$, where c is predetermined as $\frac{1}{4}$, $\frac{1}{2}$ and 1.

The results are summarized in Table 2.6 and in Table 2.7. We discover that the number of hurricanes had a significant increase around year 1933-1936, and the strength of the hurricanes had a sharp increase around the year 1923-1924. This is consistent with historical records. In history, the 1924 hurricane Cuba was the earliest officially classified Category 5 Atlantic hurricane on the Saffir-Simpson scale, and it became the strongest hurricane on record to hit the country; 1928 Okeechobee hurricane was the second recorded hurricane to reach Category 5 status on the Saffir-Simpson Hurricane Scale in the Atlantic basin after the 1924 Cuba hurricane; The 1933 Atlantic hurricane season was the second most active Atlantic hurricane season on record with 21 storms; The 1936 season was fairly active, with 17 tropical cyclones including a tropical depression. From the analysis of the shorter series, we detect that the year 2000-2001 saw an increase in the number of hurricanes. According to National Hurricane Center, the 2001 Atlantic hurricane season produced 17 tropical storms and hurricanes.

In the third study, we consider the relationship between the number of hurricanes Y, the maximum sustained winds X_1 and maximum (minimum) central pressure for data between 1851-2008 (1951-2008) X_2. We model Y as Poisson(λ), where $\theta=\log\lambda$, $p(y,\theta)=\exp\{y\theta-e^{\theta}-\log y!\}$ and use the canonical link $\theta=(1,X)'\beta$.

For the 1851-2008 data, we take the first 50 observations, and get $\hat{\beta}=(-4.99,0.01,0.006)'$.

We also estimate the bivariate mean and covariance as $\hat{\mu} = (104.8, 982.99)'$ and $\hat{\Sigma} = \begin{pmatrix} 199.96 & -20.66 \\ -20.66 & 367.56 \end{pmatrix}$. Secondly, we select $\delta = c\hat{\beta}$, where $c = \frac{1}{4}, \frac{1}{2}, 1$. Next we search for L, assuming $ARL_0 = 200$. To implement this, we simulate the bivariate series X using $\hat{\mu}$ and $\hat{\Sigma}$. Based on equation $\log(\hat{\lambda}) = (1, X)'\hat{\beta}$, we get $\hat{\lambda}$, and we can simulate Y from Poisson ($\hat{\lambda}$). Construct the CUSUM statistic and the stopping rule $S_n - \min_{0 \leq k < n} S_k \geq L$ to satisfy $ARL_0 = 200$. Finally, we fit the stopping rule to the real data and discover the signal. Results shows that there is no significant change in terms of β, which means the way how the maximum sustained winds and maximum central pressure of a hurricane relates to the number of hurricanes has not changed over the past 158 years.

For the 1951-2008 data, we take the first 20 observations, and get $\hat{\beta} = (3.08, 0.003, -0.0016)'$. We estimate the bivariate normal mean and covariance as $\hat{\mu} = \begin{pmatrix} 129.5 \\ 937.6 \end{pmatrix}$, and $\hat{\Sigma} = \begin{pmatrix} 376.05 & -220.47 \\ -220.47 & 237.41 \end{pmatrix}$. Secondly, we select $\delta = c\hat{\beta}$, where $c = \frac{1}{4}, \frac{1}{2}, 1$. Results shows that there is no significant change in terms of β, which means the way how the maximum sustained winds and minimum central pressure of a hurricane relate to the number of hurricanes has not changed over the past 58 years. Thus, the third part of our study shows broad physical relations between windspeeds and pressures have not changed, which is to be expected.

Type of Distribution	Density Function	EF-CUSUM based on
Binomial(n,p): $p \to p+\delta$	$\binom{n}{k} p^x (1-p)^{n-x}$	$xlog(\frac{p+\delta}{p}) + (N-x)log(\frac{1-p-\delta}{1-p})$
Poisson(λ): $\lambda \to \lambda+\delta$	$\frac{\lambda^x e^{-\lambda}}{\lambda!}$	$xlog\frac{\lambda+\delta}{\lambda} - \delta$
Gamma(α,β): $\alpha \to \alpha+\delta_1$, $\beta \to \beta+\delta_2$	$\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$	$\frac{\delta_2}{\beta(\beta+\delta_2)}x + \delta_1 log\frac{x}{\beta+\delta_2} - \alpha log\frac{\beta+\delta_2}{\beta} - log\frac{\Gamma(\alpha+\delta_1)}{\Gamma(\alpha)}$
Multivariate normal: $N_p(\mu,\Sigma) \to N_p(\mu+\delta,\Sigma)$ Σ is positive definite	$\frac{1}{(2\pi)^{\frac{p}{2}}\lvert\Sigma\rvert^{\frac{1}{2}}} \exp\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\}$	$(x-\mu-\frac{1}{2}\delta)'\Sigma^{-1}\delta$

Table 2.1: Exponential Family CUSUM: Binomial, Exponential, Gamma and Multivariate Normal distributions

Distribution	CUSUM statistic
$N(\mu,\sigma_1^2) \to N(\mu+\delta_1,\sigma_2^2)$	$log\sigma_1 + \frac{1}{2}\sigma_1^{-2}(x_i-\mu)^2 - log\sigma_2 - \frac{1}{2}\sigma_2^{-2}(x_i-\mu-\delta_1)^2$
$N(\mu,\sigma^2) \to N(\mu+\delta,\sigma^2)$	$\sigma^{-2}(x_i-\mu-\frac{1}{2}\delta_1)\delta_1 \propto (x_i-\mu-\frac{1}{2}\delta_1)\delta_1$
$N(\mu,\sigma_1^2) \to N(\mu,\sigma_2^2)$	$log(\sigma_2^{-1}\sigma_1) + \frac{1}{2}\sigma_1^{-2}\sigma_2^{-2}(\sigma_2^2-\sigma_1^2)(x_i-\mu)^2$
$N(\theta,\theta^2) \to N(\theta+\delta_1,(\theta+\delta_1)^2)$	$log((\theta+\delta_1)^{-1}\theta) + \frac{1}{2}\theta^{-2}(x_i-\theta)^2 - \frac{1}{2}(\theta+\delta_1)^{-2}(x_i-\theta-\delta_1)^2$

Table 2.2: CUSUM Statistic for Normal Distribution: The first row is more general with both mean and variance change. The rest three rows are special cases of the first one.

Method	Mean	Median	Std.Deviation	Max
EF-CUSUM	18.45902	15	14.36446	124
Gaussian-CUSUM	21.51974	16	17.84431	136
EF-CUSUM	9.049310	7	6.562318	68
Gaussian-CUSUM	10.678476	8	8.520224	65
EF-CUSUM	79.83406	59	71.60187	477
Gaussian-CUSUM	85.44692	62	79.26974	551

Table 2.3: Simulated binomial distribution changes: The three rows describe change from binomial(5,0.95) to binomial(5,0.90), from binomial(15,0.95) to binomial(15,0.90) and from binomial(5,0.95) to binomial(5,0.94) respectively. Here τ is fixed at 99 for illustration.

Method	Mean	Median	Std.Deviation	Max
EF-CUSUM	101.51383	74	95.09682	730
Gaussian-CUSUM	111.9772	80	106.36463	755
EF-CUSUM	93.05656	68	86.17123	546
Gaussian-CUSUM	98.06841	71	90.09476	634
EF-CUSUM	4.480411	4	2.815292	22
Gaussian-CUSUM	4.587708	4	3.149742	22
EF-CUSUM	2.858086	3	1.232671	11
Gaussian-CUSUM	3.085586	3	1.159972	11

Table 2.4: Simulated Poisson distribution changes: The four rows describe change from Poisson(3) to Poisson(3.1), from Poisson(3) to Poisson(2.9), from Poisson(4) to Poisson(7) and from Poisson(4) to Poisson(1) respectively. Here τ is fixed at 99 for illustration.

Method	Mean	Median	Std.Deviation	Max
EF-CUSUM	9.923995	8	6.839445	53
Gaussian-CUSUM	15.64539	12	12.99359	165
EF-CUSUM	28.95583	25	19.30729	152
Gaussian-CUSUM	35.11858	29	27.01831	248
EF-CUSUM	70.26987	55	58.85137	437
Gaussian-CUSUM	75.67127	60	61.48744	417
EF-CUSUM	16.24952	13	11.87364	130
Gaussian-CUSUM	21.89900	17	18.15123	142
EF-CUSUM	1.063716	1	0.2497610	3
Gaussian-CUSUM	1.069783	1	0.2795194	3

Table 2.5: Simulated Gamma distribution changes: The five rows describe change from Gamma(1,2) to Gamma(1.5,2.5), from Gamma(1,2) to Gamma(1.5,2.5), from Gamma(3,4) to Gamma(3.5,3.5), from Gamma(3,4) to Gamma(3.5,4.5) and from Gamma(10,10) to Gamma(17,18) respectively. Here τ is fixed at 99 for illustration.

Distribution	$c = \frac{1}{4}$	c=$\frac{1}{2}$	c=1
Poisson	1936	1933	1933
Bivariate Normal	1924	1923	1924

Table 2.6: Atlantic tropical storm data from 1851 to 2008 are used to detect any mean change in tropical storm characteristics. Here c is the magnitude representing the number of standard deviation from the mean. Result shows that the number of tropical storm had a significant increase around 1933-1936, and strength of the tropical storm increased around 1923-1924.

Distribution	$c = \frac{1}{4}$	c=$\frac{1}{2}$	c=1
Poisson	2001	2001	2000
Bivariate Normal	2008	2008	2008

Table 2.7: Atlantic tropical storm data from 1951 to 2008 are used to detect any mean change in tropical storm characteristics. Here c is the magnitude representing the number of standard deviation from the mean. Result shows that the number of tropical storm had a significant increase around the year of 2000, and strength of the tropical storm has not changed.

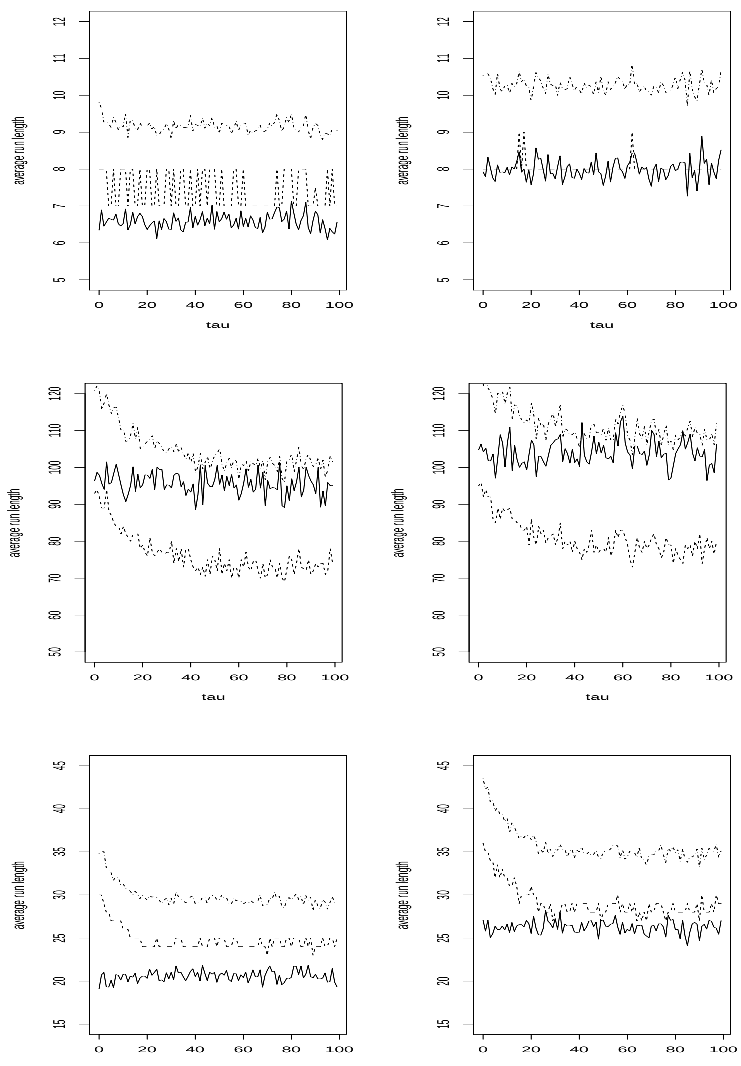


Figure 2.1: Performance Comparison: Exponential Family CUSUM with Normal CUSUM. Dotdash, dashed and solid line stand for mean, median and standard deviation. The top panel describes run length comparison from Binomial(15,0.95) to Binomial(15,0.90), the middle panel describes run length comparison from Poisson(3) to Poisson(3.1), the bottom panel describes run length comparison from Gamma(1,2) to Gamma(1.5,1.5). Due to length limitation of the graphs, we here do not include the MAX line.

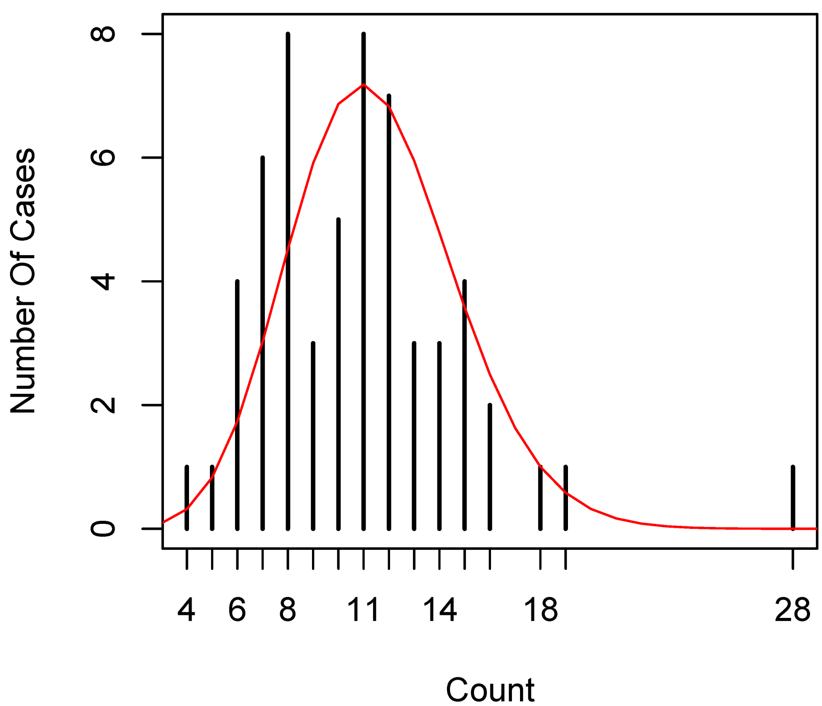


Figure 2.2: The observed data and a Poisson fit for the yearly number of tropical storms between the years 1951-2008.

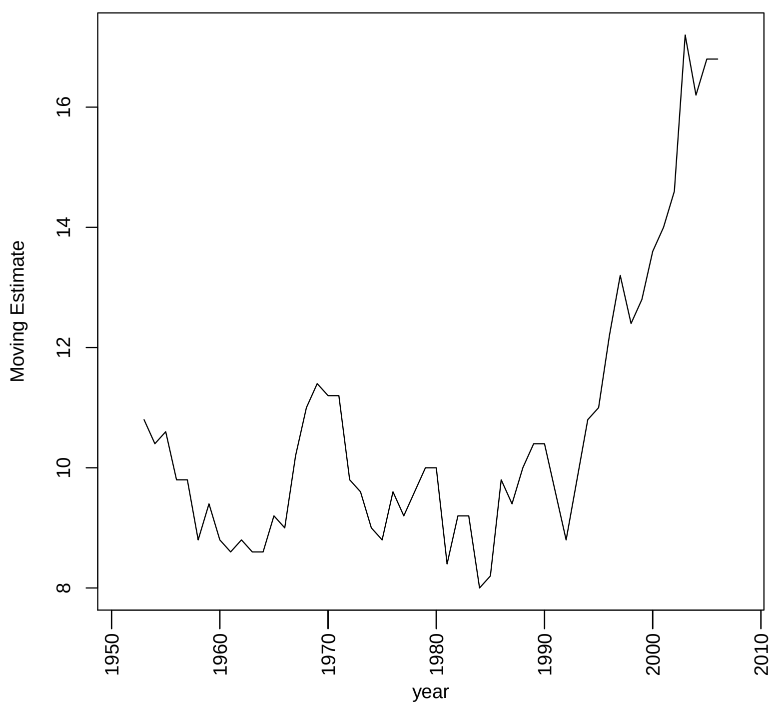


Figure 2.3: A moving average estimate of the yearly number of tropical storms between the years 1951-2008.

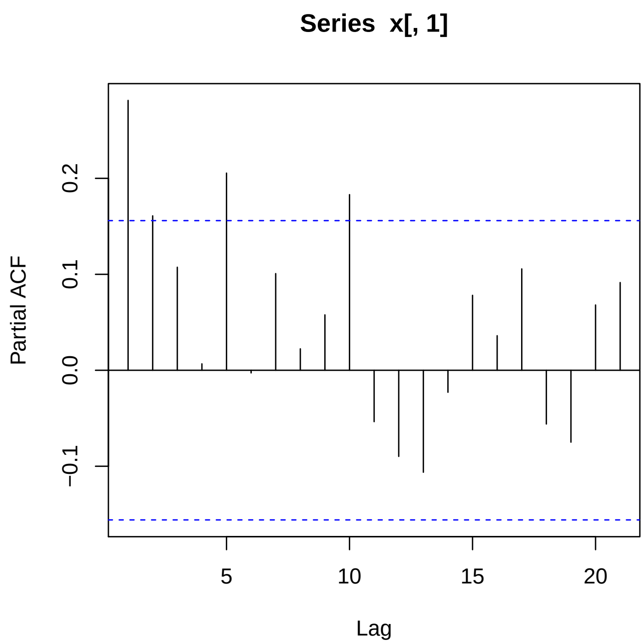


Figure 2.4: PACF function of the yearly number of tropical storms between the years 1951-2008.

Chapter 3

Detecting Change in the Extremes

The generalized extreme value family distribution not only plays an important role in real data analysis, but also represents a non-regular family of distributions in which its support depends on the parameter. This chapter further develops the idea of Chapter 2 and discusses the change detection problem in the extreme data by examining the maxima (minima). As a motivation example, in Section 3.1, we study the change from Uniform $[0, \theta]$ to Uniform $[0, \theta + \delta]$ where $\delta > 0$ and θ, δ are known. The Gumbel distribution, however, has a regular analytical density function; therefore the detecting procedure is similar to the exponential family, in which a CUSUM type of statistic is derived, as discussed by [112], [113] and [14]. However, when it comes to Fréchet distribution and Weibull distribution, the decision rule exhibits a variety of forms, which may be a combination of the CUSUM type statistic and the simple judgement of the $X_j's$ over a threshold. Researchers such as [78], [89], [153], [155] and [156] discussed the GEV distribution. In Section 3.2, we present a systematic GEV based likelihood ratio procedure, with Section 3.5.1, Section 3.2.2 and Section 3.2.3 covering each type of the GEV distribution. The Generalized Pareto distribution is treated in Section 3.3. Section 3.5 illustrates the advantage of the GEV likelihood based procedure over the normality based procedure with both ARL and p-value criteria. Finally, the changing behavior of the maximum sustained wind speeds is studied in Section 3.6.

3.1 Motivation Example: Uniform Distribution

In this section, we consider the uniform distribution $[0, \theta]$, which is one of the simplest probability distributions whose support depends on the parameter. Denote $\overset{\text{iid}}{\sim}$ to be independently and identically distributed. Assume $X_1, \ldots, X_\tau$ are $\overset{\text{iid}}{\sim}$ uniformly distributed on the interval $[0, \theta]$ and $X_{\tau+1}, \ldots, X_n$.. are $\overset{\text{iid}}{\sim}$ uniformly distributed on the interval $[0, \theta + \delta]$, where θ, δ are known or unknown, but τ is an unknown parameter. Our job is to develop an algorithm to signal an alarm when a change occurs, and not to signal an alarm when there is no change. We discuss for $\delta > 0$ and $\delta < 0$ within the hypothesis testing framework. The level of test is denoted as α, where $0 < \alpha < 1$.

3.1.1 Known Parameter Case

$\underline{\delta > 0}$.

Theorem 3.1.1. *For the hypothesis $H_0 : \tau \geq n$ Vs $H_1 : 0 \leq \tau < n$, if there exists j, $1 \leq j \leq n$, such that $X_j > \theta$, then H_1 holds; otherwise H_0 holds. As a convention, $\infty \times 0 = 0$.*

Proof. Under H_0, the likelihood function is

$$L_0(\tau) = \prod_{i=1}^{n} \theta^{-1} I_{[0,\theta]}(X_i).$$

Under H_1, the likelihood function is

$$L_1(\tau) = \prod_{i=1}^{\tau} \theta^{-1} I_{[0,\theta]}(X_i) \prod_{i=\tau+1}^{n} (\theta + \delta)^{-1} I_{[0,\theta+\delta]}(X_i)$$

If $\exists j$, $1 \leq j \leq n$, such that $X_j > \theta$, then $L_0(\tau) = 0$. We reject H_0 and conclude a change.

Otherwise, if $\forall j$, $X_j \leq \theta$, then the likelihood ratio statistic is

$$\begin{aligned}
\Lambda(\tau) &= \frac{\max_{0\leq\tau<n} L_1(\tau)}{\max_{\tau\geq n} L_0(\tau)} \\
&\propto \max_{0\leq\tau<n} \prod_{i=1}^{\tau} \theta^{-1} I_{[0,\theta]}(X_i) \prod_{i=\tau+1}^{n} (\theta + \delta)^{-1} I_{[0,\theta+\delta]}(X_i) \\
&= \max_{0\leq\tau<n} \theta^{-\tau} (\theta + \delta)^{-(n-\tau)} \\
&= \max_{0\leq\tau<n} \{\theta^{-1}(\theta + \delta)\}^{\tau} (\theta + \delta)^{-n}
\end{aligned}$$

Since $\theta^{-1}(\theta+\delta) > 1$, $\hat{\tau} = n-1$ will maximize the above. In such case,

$$\begin{aligned}\Lambda(\hat{\tau}) &= \{\prod_{i=1}^{n} \theta^{-1}\}^{-1}\{\prod_{i=1}^{n-1} \theta^{-1}\}(\theta+\delta)^{-1} \\ &= (\theta+\delta)^{-1}\theta\end{aligned}$$

Therefore,

$$\Lambda(\hat{\tau}) = \infty I_{\{\exists j, 1\le j\le n, s.t. X_j > \theta\}} + (\theta+\delta)^{-1}\theta I_{\{\forall j, X_j \le \theta\}}$$

The rejection region is $\Lambda(\hat{\tau}) \geq C$ for some constant C. Under the null, to satisfy the level α condition, we ask $P_0(\Lambda(\hat{\tau}) \geq C) \leq \alpha$. If $C \leq (\theta+\delta)^{-1}\theta$, then $P_0(\Lambda(\hat{\tau}) \geq C) = 1 > \alpha$, which contradicts the assumption of $\alpha < 1$. Therefore $C > (\theta+\delta)^{-1}\theta$. Interestingly in this case, the rejection region is an empty set, which literally means we will never reject. □

Remark This example gives us two interesting thoughts. Firstly, we know that the null hypothesis is protected in the hypothesis testing procedure, so it is natural that when $X_j \leq \theta$, we are not in favor of $[0, \theta+\delta]$ at all. But when one of the $X_j's$ are above the threshold θ, we know immediately that $[0, \theta]$ can not be true.
Secondly, The power of this test is computed as follows:

$$P_1(\Lambda(\hat{\tau}) \geq C|H_1) = P_1(\exists j, 1 \leq j \leq n, s.t. X_j > \theta) = 1 - P_1(\forall j, X_j \leq \theta) = 1 - ((\theta+\delta)^{-1}\theta)^{n-\tau} \to 1$$

as $n \to \infty$ and $\tau = o(n)$. This is again intuitive because as more data come in, the chance we make errors should shrink towards zero.

Corollary 3.1.2. *$ARL_0 = \infty$, $ARL_1 = \delta^{-1}(\theta+\delta)$.*

$\underline{\delta < 0}$.

Theorem 3.1.3. *For the hypothesis testing, $H_0 : \tau \geq n$ Vs $H_1 : 0 \leq \tau < n$, Define $k_0 = \inf\{t : X_{t+1}, \ldots, X_n \leq \theta+\delta\}$, and $\infty = \inf\{\emptyset\}$ by convention. The rejection region is $\{n - k_0 \geq [\{\ln(\theta^{-1}(\theta+\delta))\}^{-1}\ln(\alpha)] + 1\}$ for the level α test.*

Proof. Obviously, τ should be at least k_0.

$$\begin{aligned}\Lambda(\tau) &= \frac{\max_{k_0\le\tau<n} L_1(\tau)}{\max_{\tau\ge n} L_0(\tau)} \\ &\propto \max_{k_0\le\tau<n} \prod_{i=1}^{\tau}\theta^{-1}\prod_{i=\tau+1}^{n}(\theta+\delta)^{-1} \\ &= \max_{k_0\le\tau<n} \theta^{-\tau}(\theta+\delta)^{-(n-\tau)} \\ &= \max_{k_0\le\tau<n} \{\theta^{-1}(\theta+\delta)\}^{\tau}(\theta+\delta)^{-n}\end{aligned}$$

Since $0<\theta^{-1}(\theta+\delta)<1$, $\hat{\tau}$ should be chosen as k_0 to maximize $\Lambda(\tau)$. So the likelihood ratio statistic is

$$\begin{aligned}\Lambda(\hat{\tau}) &= \{\prod_{i=1}^{n}\theta^{-1}\}^{-1}\prod_{i=1}^{k_0}\theta^{-1}\prod_{i=k_0+1}^{n}(\theta+\delta)^{-1} \\ &= \{(\theta+\delta)^{-1}\theta\}^{n-k_0}\end{aligned}$$

The rejection region is $\left\{(\theta+\delta)^{-1}\theta\right\}^{n-k_0}\ge L$ for some constant L, which is equivalent to $n-k_0\ge L$. So $P_0(n-k_0\ge L)=P_0(X_{k_0+1}\le\theta+\delta,\dots,X_n\le\theta+\delta)\le\left\{\theta^{-1}(\theta+\delta)\right\}^{L}\le\alpha$. Choose $L=[\left\{\ln(\theta^{-1}(\theta+\delta))\right\}^{-1}\ln(\alpha)]+1$, where $[x]$ is defined as the greatest integer function. We reject when $\{n-k_0\ge[\left\{\ln(\theta^{-1}(\theta+\delta))\right\}^{-1}\ln(\alpha)]\}$. ☐

Remark When all $X_n>\theta+\delta$, the rejection region becomes $-\infty\ge[\left\{\ln(\theta^{-1}(\theta+\delta))\right\}^{-1}\ln(\alpha)]+1$, which can never happen, and we are not able to reject. This is again consistent with the philosophy of protecting the null hypothesis.

Remark In the above situation, assume τ is sufficiently large, $(\tau>L)$, then

$$ARL_1\le\sum_{i=1}^{L}i(\theta^{-1}(\theta+\delta))^{L-i}(-\theta^{-1}\delta)=L+(1-(\theta^{-1}(\theta+\delta))^{L})\delta^{-1}(\theta+\delta).$$

3.1.2 Unknown Parameter Case

In this section, we treat $\theta>0$ and δ as unknown parameters, and the change point τ unknown as well. Our null hypothesis is $H_0:\tau\ge n$ and the alternative $0\le\tau<n$.

The likelihood function under the null is

$$L_0(\tau,\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I_{[0,\theta]}(X_i) \tag{3.1.1}$$

The likelihood under the alternative is

$$L_1(\tau,\theta,\delta) = \prod_{i=1}^{\tau} \frac{1}{\theta} I_{[0,\theta]}(X_i) \prod_{i=\tau+1}^{n} \frac{1}{\theta+\delta} I_{[0,\theta+\delta]}(X_i) \tag{3.1.2}$$

To maximize L_0 over τ and θ, note that L_0 does not depend on τ, so we only need to minimize over θ.

$$L_0(\tau,\theta) = (\frac{1}{\theta})^n I_{[0,\theta]}(\max_{1\le i\le n} X_i) \tag{3.1.3}$$

To maximize L_0, we need θ to be as small as possible, but not smaller than $\max_{1\le i\le n} X_i$. Therefore, $\hat{\theta} = \max_{1\le i\le n} X_i$.

$$L_0 = (\frac{1}{\max_{1\le i\le n} X_i})^n \tag{3.1.4}$$

To simplify L_1,

$$L_1(\tau,\theta,\delta) = (\frac{1}{\theta})^\tau (\frac{1}{\theta+\delta})^{n-\tau} I_{[0,\theta]}(\max_{1\le i\le \tau} X_i) I_{[0,\theta+\delta]}(\max_{\tau+1\le i\le n} X_i) \tag{3.1.5}$$

If we fix τ, it is easy to see $\hat{\theta} = \max_{1\le i\le \tau} X_i$, and $\theta+\delta = \max_{\tau+1\le i\le n} X_i$.

$$L_1(\tau,\hat{\theta},\hat{\delta}) = (\frac{1}{\max_{1\le i\le \tau} X_i})^\tau (\frac{1}{\max_{\tau+1\le i\le n} X_i})^{n-\tau} \tag{3.1.6}$$

Denote $\hat{\tau}$ to be the maximizer. Therefore, the likelihood ratio becomes

$$\Lambda = \frac{L_1(\hat{\tau})}{L_0} \tag{3.1.7}$$

We reject when $\lambda \ge L$.

To determine L, we need bootstrap along with the following theorem:

Theorem 3.1.4. *For $\{X_i i = 1, 2, \ldots, n\}$ independently sampled from $U[0, \theta]$, where θ is a fixed but unknown parameter, $\max_{1 \le i \le n} X_i \to \theta$ with probability 1, and therefore almost surely.*

Proof.

$$P(|\max_{1 \le i \le n} X_i - \theta| > \epsilon) = P(\max_{1 \le i \le n} X_i < \theta - \epsilon) = (\frac{\theta - \epsilon}{\theta})^n \to 0 \tag{3.1.8}$$

as $n \to \infty$. Since θ is a constant, convergence in probability implies convergence almost surely. □

By assuming we have the first k observations to follow $U[0, \theta]$, we have a consistent estimate of θ, which is $\hat{\theta} = \max_{1 \le i \le k} X_i$. Then we simulate large enough samples according to the uniform distribution $U[0, \hat{\theta}]$, and choose L according to predefined ARL_0 defined in Chapter 2 to control for false alarm. We also note that the estimator $\hat{\tau}$ can not be consistent in this case because the maxima both before and after the change appears in any position with equal probability in the respective phase of observation.

3.2 GEV Likelihood Based Procedure

The generalized extreme value distribution can be classified into three categories: Gumbel, Fréchet and reversed Weibull distribution, which are known as type I, type II and type III generalized extreme value distributions. In the following paragraph, we discuss each type of the generalized extreme value distribution and summarize their detection algorithms.

3.2.1 Gumbel Distribution

The Gumbel distribution has its cumulative form:

$$F(X; \mu, \sigma) = \exp\left\{-\exp\left\{-\sigma^{-1}(X - \mu)\right\}\right\}$$

for any $X \in R$. The density function can be easily derived as

$$f(X; \mu, \sigma) = \sigma^{-1} \exp\left\{-\sigma^{-1}(X - \mu)\right\} \exp\left\{-\exp\left\{-\sigma^{-1}(X - \mu)\right\}\right\}$$

where $X \in R$ and $\sigma > 0$.

Let τ be an unknown parameter. Assume that $X_1, \ldots, X_\tau$ are $\overset{\text{iid}}{\sim}$ with Gumbel density function $f(X; \mu, \sigma_1)$, $X_{\tau+1}, \ldots, X_n, ...$ are $\overset{\text{iid}}{\sim}$ with Gumbel density function $f(X; \mu + \delta, \sigma_2)$.

Theorem 3.2.1. *For the hypothesis testing problem,*

$$H_0 : \tau \geq n.$$
$$H_1 : 0 \leq \tau < n.$$

Denote $Y_i = \log \sigma_1 + \sigma_1^{-1}(X_i - \mu) + \exp\{-\sigma_1^{-1}(X_i - \mu)\} - \log \sigma_2 - \sigma_2^{-1}(X_i - \mu - \delta) - \exp\{-\sigma_2^{-1}(X_i - \mu - \delta)\}$, *and* $S_k = \sum_{i=1}^{k} Y_i$, *we reject if* $S_n - \min_{0 \leq k < n} S_k \geq L$ *for some constant* L.

Proof. Since Gumbel distribution is a regular distribution, we may follow the procedure in [14] easily. □

3.2.2 Fréchet Distribution

The cumulative Fréchet distribution is expressed as

$$F(x; \mu, \sigma, \alpha) = \exp\left\{-(\sigma^{-1}(x - \mu))^{-\alpha}\right\} I_{(\mu,\infty)(x)}.$$

By taking the derivative of $F(x; \mu, \sigma, \alpha)$, the density function becomes

$$\begin{cases} \sigma^{-1}\alpha(\sigma^{-1}(x-\mu))^{-\alpha-1} \exp\left\{-(\sigma^{-1}(x-\mu))^{-\alpha}\right\} & x > \mu \\ \lim\limits_{x \to \mu} (x-\mu)^{-1} \exp\left\{-(\sigma^{-1}(x-\mu))^{-\alpha}\right\} = 0 & x = \mu \\ 0 & x < \mu \end{cases}$$

Therefore the density function can be written in the following form

$$f(x; \mu, \sigma, \alpha) = \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(x-\mu))^{-\alpha}\right\} (\sigma^{-1}(x-\mu))^{-\alpha-1} I_{(\mu,\infty)}(x)$$

In this paper, we list three cases.

$\underline{\delta > 0.}$ Assume $X_1, \ldots, X_\tau$ coming from $\overset{\text{iid}}{\sim}$ with Fréchet density function $f(x; \mu, \sigma, \alpha)$, $X_{\tau+1}, \ldots, X_n$ coming from $\overset{\text{iid}}{\sim}$ with Fréchet density function $f(x; \mu + \delta, \sigma, \alpha)$.

Theorem 3.2.2. *For the hypothesis testing problem*

$$H_0 : \tau \geq n.$$
$$H_1 : 0 \leq \tau < n.$$

Denote $k_0 = \inf\{t : X_{t+1}, \ldots, X_n > \mu + \delta\}$, $Y_i = (\sigma^{-1}(X_i - \mu))^{-\alpha} - (\sigma^{-1}(X_i - \mu - \delta))^{-\alpha} + (\alpha+1)\log(X_i - \mu) - (\alpha+1)\log(X_i - \mu - \delta)$ *and* $\tilde{S}_k = \sum_{i=k+1}^{n} Y_i$, *we reject if* $\max_{k_0 \leq k < n} \tilde{S}_k \geq L$. *In particular, if* $\{t : X_{t+1}, \ldots, X_n > \mu + \delta\} = \varnothing$, $k_0 = \infty$ *and by convention,* $\max_{k_0 \leq k < n} \tilde{S}_k = 0$, *indicating no rejection.*

Proof. The likelihood function for the null hypothesis is

$$L_0(\tau) = \prod_{i=1}^{n} \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(X_i - \mu))^{-\alpha}\right\}(\sigma^{-1}(X_i - \mu))^{-\alpha-1} I_{(\mu,\infty)}(X_i)$$

The likelihood function for the alternative hypothesis is

$$\begin{aligned} L_1(\tau) &= \prod_{i=1}^{\tau} \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(X_i - \mu))^{-\alpha}\right\}(\sigma^{-1}(X_i - \mu))^{-\alpha-1} I_{(\mu,\infty)}(X_i) \\ &\times \prod_{i=\tau+1}^{n} \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(X_i - \mu - \delta))^{-\alpha}\right\}(\sigma^{-1}(X_i - \mu - \delta))^{-\alpha-1} I_{(\mu+\delta,\infty)}(X_i) \end{aligned}$$

The likelihood ratio statistic is

$$\begin{aligned} \Lambda(\tau) &= \frac{\max_{0 \leq \tau < n} L_1(\tau)}{\max_{\tau \geq n} L_0(\tau)} \\ &\propto \max_{0 \leq \tau < n} \prod_{i=1}^{\tau} \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(X_i - \mu))^{-\alpha}\right\}(\sigma^{-1}(X_i - \mu))^{-\alpha-1} I_{(\mu,\infty)}(X_i) \\ &\times \prod_{i=\tau+1}^{n} \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(X_i - \mu - \delta))^{-\alpha}\right\}(\sigma^{-1}(X_i - \mu - \delta))^{-\alpha-1} I_{(\mu+\delta,\infty)}(X_i) \end{aligned}$$

Maximizing Λ is equivalent to maximizing L_1 over τ, and τ must satisfy the following for $L_1(\tau)$ to be positive.

$$\tau = \{t : X_{t+1}, \ldots, X_n > \mu + \delta\} \tag{3.2.9}$$

Here we define $k_0 = \inf\{t : X_{t+1}, \dots, X_n > \mu + \delta\}$. Under (3.2.9),we wish to maximize

$$\begin{aligned} & \prod_{i=1}^{\tau} \exp\left\{-(\sigma^{-1}(X_i-\mu))^{-\alpha}\right\}(\sigma^{-1}(X_i-\mu))^{-\alpha-1} \\ \times & \prod_{i=\tau+1}^{n} \exp\left\{-(\sigma^{-1}(X_i-\mu-\delta))^{-\alpha}\right\}(\sigma^{-1}(X_i-\mu-\delta))^{-\alpha-1} \end{aligned}$$

This is equivalent to maximizing

$$\begin{aligned} & \sum_{i=1}^{\tau}\left\{-(\sigma^{-1}(X_i-\mu))^{-\alpha}-(\alpha+1)\log(\sigma^{-1}(X_i-\mu))\right\} \\ + & \sum_{i=\tau+1}^{n}\left\{-(\sigma^{-1}(X_i-\mu-\delta))^{-\alpha}-(\alpha+1)\log(\sigma^{-1}(X_i-\mu-\delta))\right\} \\ = & \sum_{i=1}^{n}\left\{-(\sigma^{-1}(X_i-\mu))^{-\alpha}-(\alpha+1)\log(\sigma^{-1}(X_i-\mu))\right\} \\ + & \sum_{i=\tau+1}^{n}\{(\sigma^{-1}(X_i-\mu))^{-\alpha}+(\alpha+1)\log(\sigma^{-1}(X_i-\mu)) \\ - & (\sigma^{-1}(X_i-\mu-\delta))^{-\alpha}-(\alpha+1)\log(\sigma^{-1}(X_i-\mu-\delta))\} \end{aligned}$$

If we let $Y_i = (\sigma^{-1}(X_i-\mu))^{-\alpha} - (\sigma^{-1}(X_i-\mu-\delta))^{-\alpha} + (\alpha+1)\log(X_i-\mu) - (\alpha+1)\log(X_i-\mu-\delta)$, we should choose $\hat{\tau}$ such that $\tilde{S}_k = \sum_{i=k+1}^{n} Y_i$ is maximized. From (3.2.9), τ has to be at least k_0. Therefore $\hat{\tau} = \arg\max_{k_0 \le k < n} \tilde{S}_k$.
Thus

$$\begin{aligned} \Lambda(\tau) &= \exp\{\sum_{i=\hat{\tau}+1}^{n}\{(\sigma^{-1}(X_i-\mu))^{-\alpha}-(\sigma^{-1}(X_i-\mu-\delta))^{-\alpha} \\ &+ (\alpha+1)\log(X_i-\mu)-(\alpha+1)\log(X_i-\mu-\delta)\}\} \\ &= \exp\left\{\sum_{i=\hat{\tau}+1}^{n} y_i\right\} \\ &= \exp\left\{\max_{k_0\le k<n}\tilde{S}_k\right\} \end{aligned}$$

Therefore, we reject the null hypothesis if $\max_{k_0 \le k < n} \tilde{S}_k \ge L$ for some constant L. □

<u>$\delta < 0$.</u> Assume $X_1, \dots, X_\tau$ coming from $\stackrel{\text{iid}}{\sim}$ with Fréchet density function $f(x; \mu, \sigma, \alpha)$, $X_{\tau+1}, \dots, X_n$ coming from $\stackrel{\text{iid}}{\sim}$ with Fréchet density function $f(x; \mu + \delta, \sigma, \alpha)$

Theorem 3.2.3. *For the hypothesis testing problem:*

$$H_0 : \tau \geq n.$$
$$H_1 : 0 \leq \tau < n.$$

Define $Y_i = (\sigma^{-1}(X_i-\mu))^{-\alpha} - (\sigma^{-1}(X_i-\mu-\delta))^{-\alpha} + (\alpha+1)\log(X_i-\mu) - (\alpha+1)\log(X_i-\mu-\delta)$, *and* $S_k = \sum_{i=1}^{k} Y_i$. *We reject either if one of* $X_j \leq \mu$ *or if* $S_n - \min_{0\leq k<n} S_k \geq L$ *under the condition that* $X_j > \mu$ *for all* j.

Remark When $\delta < 0$, Readers might be curious what would happen under the condition of $X_i > \mu$. Intuitively, X_i should be more likely from $f(x;\mu,\sigma,\alpha)$ than $f(x;\mu+\delta,\sigma,\alpha)$ when $X_i > \mu$. To investigate, we study how

$$f(x;\mu,\sigma,\alpha) = \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(x-\mu))^{-\alpha}\right\}(\sigma^{-1}(x-\mu))^{-\alpha-1} I_{(\mu,\infty)}(x)$$

responds with μ when $x > \mu$. This is equivalent to comparing two density functions: $f(x;\mu,\sigma,\alpha|x>\mu)$ and $f(x;\mu+\delta,\sigma,\alpha|x>\mu)$.
Define $g(\mu) = \exp\left\{-(\sigma^{-1}(x-\mu))^{-\alpha}\right\}(\sigma^{-1}(x-\mu))^{-\alpha-1}$.

$$\frac{\partial g}{\partial \mu} = \sigma^{-1}\exp\left\{-(\sigma^{-1}(x-\mu))^{-\alpha}\right\}(\sigma^{-1}(x-\mu))^{-2\alpha-2}\left\{(\alpha+1)(\sigma^{-1}(x-\mu))^{\alpha} - \alpha\right\}$$

When $\sigma^{-1}(x-\mu) > \left\{(\alpha+1)^{-1}\alpha\right\}^{\frac{1}{\alpha}}$, $g(\mu)$ is an increasing function with μ, so a sufficient condition for $g(\mu) > g(\mu+\delta)$ is that

$$\sigma^{-1}(x-\mu) > \left\{(\alpha+1)^{-1}\alpha\right\}^{\frac{1}{\alpha}} \tag{3.2.10}$$

There are two interesting cases to consider:
1. α is small.

$$\lim_{\alpha\to 0}\left\{(\alpha+1)^{-1}\alpha\right\}^{\frac{1}{\alpha}} = 0$$

2. α is large.

$$\lim_{\alpha\to \infty}\left\{(\alpha+1)^{-1}\alpha\right\}^{\frac{1}{\alpha}} = 1$$

By (3.2.10), if $X_i > \mu$, in the case of α being small, X_i is more likely to be from the first distribution; in the case of α being large, we need to see how large X_i is to determine which distribution X_i is more likely to come from.

Change in α. Assume $X_1, \ldots, X_\tau$ coming from $\overset{\text{iid}}{\sim}$ with Fréchet density function $f(x;\mu,\sigma,\alpha_1)$, and $X_{\tau+1}, \ldots, X_n$ coming from $\overset{\text{iid}}{\sim}$ with $f(x;\mu,\sigma,\alpha_2)$.

Theorem 3.2.4. *For the hypothesis testing problem*

$$H_0 : \tau \geq n.$$
$$H_1 : 0 \leq \tau < n.$$

define $Y_i = \log(\alpha_1^{-1}\alpha_2) + (\alpha_1 - \alpha_2)\log(\sigma^{-1}(X_i - \mu)) + (\sigma^{-1}(X_i - \mu))^{-\alpha_1} - (\sigma^{-1}(X_i - \mu))^{-\alpha_2}$, *and* $S_k = \sum_{i=1}^{k} y_i$, *we reject when* $S_n - \min_{0 \leq k < n} S_k \geq L$ *for some constant* L.

Proof. The likelihood function for the null hypothesis is

$$L_0(\tau) = \prod_{i=1}^{n} \sigma^{-1}\alpha_1 \exp\left\{-(\sigma^{-1}(X_i - \mu))^{-\alpha_1}\right\}(\sigma^{-1}(X_i - \mu))^{-\alpha_1 - 1} I_{(\mu,\infty)}(X_i)$$

The likelihood function for the alternative hypothesis is

$$\begin{aligned} L_1(\tau) &= \prod_{i=1}^{\tau} \sigma^{-1}\alpha_1 \exp\left\{-(\sigma^{-1}(X_i - \mu))^{-\alpha_1}\right\}(\sigma^{-1}(X_i - \mu))^{-\alpha_1 - 1} I_{(\mu,\infty)}(X_i) \\ &\times \prod_{i=\tau+1}^{n} \sigma^{-1}\alpha_2 \exp\left\{-(\sigma^{-1}(X_i - \mu))^{-\alpha_2}\right\}(\sigma^{-1}(X_i - \mu))^{-\alpha_2 - 1} I_{(\mu,\infty)}(X_i) \end{aligned}$$

The likelihood ratio statistic is

$$\begin{aligned} \Lambda(\tau) = \frac{\max_{0 \leq \tau < n} L_1(\tau)}{\max_{\tau \geq n} L_0(\tau)} &\propto \max_{0 \leq \tau < n} \prod_{i=1}^{\tau} \sigma^{-1}\alpha_1 \exp\left\{-(\sigma^{-1}(X_i - \mu))^{-\alpha_1}\right\}(\sigma^{-1}(X_i - \mu))^{-\alpha_1 - 1} I_{(\mu,\infty)}(X_i) \\ &\times \prod_{i=\tau+1}^{n} \sigma^{-1}\alpha_2 \exp\left\{-(\sigma^{-1}(X_i - \mu))^{-\alpha_2}\right\}(\sigma^{-1}(X_i - \mu))^{-\alpha_2 - 1} I_{(\mu,\infty)}(X_i) \\ &\propto \max_{0 \leq \tau < n} \prod_{i=1}^{\tau} \alpha_1 \exp\left\{-(\sigma^{-1}(X_i - \mu))^{-\alpha_1}\right\}(\sigma^{-1}(X_i - \mu))^{-\alpha_1} \\ &\times \prod_{i=\tau+1}^{n} \alpha_2 \exp\left\{-(\sigma^{-1}(X_i - \mu))^{-\alpha_2}\right\}(\sigma^{-1}(X_i - \mu))^{-\alpha_2} \end{aligned}$$

Maximizing $\Lambda(\tau)$ is equivalent to maximizing

$$\begin{aligned}
& \sum_{i=1}^{\tau}\left\{\log\alpha_1-(\sigma^{-1}(X_i-\mu))^{-\alpha_1}-\alpha_1\log(\sigma^{-1}(X_i-\mu))\right\} \\
+ & \sum_{i=\tau+1}^{n}\left\{\log\alpha_2-(\sigma^{-1}(X_i-\mu))^{-\alpha_2}-\alpha_2\log(\sigma^{-1}(X_i-\mu))\right\} \\
= & -\sum_{i=1}^{\tau}\left\{\log(\alpha_1^{-1}\alpha_2)+(\alpha_1-\alpha_2)\log(\sigma^{-1}(X_i-\mu))+(\sigma^{-1}(X_i-\mu))^{-\alpha_1}-(\sigma^{-1}(X_i-\mu))^{-\alpha_2}\right\} \\
+ & \sum_{i=1}^{n}\left\{\log\alpha_2-(\sigma^{-1}(X_i-\mu))^{-\alpha_2}-\alpha_2\log(\sigma^{-1}(X_i-\mu))\right\}
\end{aligned}$$

If we define $Y_i=\log(\alpha_1^{-1}\alpha_2)+(\alpha_1-\alpha_2)\log(\sigma^{-1}(X_i-\mu))+(\sigma^{-1}(X_i-\mu))^{-\alpha_1}-(\sigma^{-1}(X_i-\mu))^{-\alpha_2}$, and $S_k=\sum_{i=1}^{k}Y_i$, we choose $\hat{\tau}=\arg\min_{0\le k<n}S_k$.

$$\begin{aligned}
\Lambda(\hat{\tau}) &= \{\prod_{i=\hat{\tau}+1}^{n}\sigma^{-1}\alpha_1\exp\left\{-(\sigma^{-1}(X_i-\mu))^{-\alpha_1}\right\}(\sigma^{-1}(X_i-\mu))^{-\alpha_1-1}I_{(\mu,\infty)}(X_i)\}^{-1} \\
&\times \prod_{i=\hat{\tau}+1}^{n}\sigma^{-1}\alpha_2\exp\left\{-(\sigma^{-1}(X_i-\mu))^{-\alpha_2}\right\}(\sigma^{-1}(X_i-\mu))^{-\alpha_2-1}I_{(\mu,\infty)}(X_i) \\
&= \exp\left\{\sum_{i=\hat{\tau}+1}^{n}Y_i\right\}=\exp\left\{S_n-\min_{0\le k<n}S_k\right\}
\end{aligned}$$

So the rejection region is $S_n-\min_{0\le k<n}S_k\ge L$ for some constant L. □

3.2.3 Reversed Weibull Distribution

The cumulative distribution of reversed Weibull distribution is

$$F(x;\mu,\delta,\alpha)=\exp\left\{-(-\sigma^{-1}(x-\mu))^{\alpha}\right\}I_{(-\infty,\mu)}(x)+I_{[\mu,\infty)}(x)$$

When $\alpha>1$, The probability density function exists, and it is

$$\begin{cases}
\sigma^{-1}\alpha(-\sigma^{-1}(x-\mu))^{\alpha-1}\exp\left\{-(-\sigma^{-1}(x-\mu))^{\alpha}\right\} & x<\mu \\
\lim_{x\to\mu^-}(\mu-x)^{-1}\left\{1-\exp\left\{-(\sigma^{-1}(\mu-x))^{\alpha}\right\}\right\}=0 & x=\mu \\
0 & x>\mu
\end{cases}$$

This density function can be rewritten in the following form:

$$f(x;\mu,\sigma,\alpha)=\sigma^{-1}\alpha(-\sigma^{-1}(x-\mu))^{\alpha-1}\exp\left\{-(-\sigma^{-1}(x-\mu))^{\alpha}\right\}I_{(-\infty,\mu)}(x)$$

When $\alpha=1$, the density function is discontinuous at μ.

$$\lim_{x\to\mu^-}(\mu-x)^{-1}\left\{1-\exp\left\{-(\sigma^{-1}(\mu-x))\right\}\right\}=\sigma^{-1}\neq 0.$$

When $\alpha<1$, the density function does not exist at μ.

$$\lim_{x\to\mu^-}(\mu-x)^{-1}\left\{1-\exp\left\{-(\sigma^{-1}(\mu-x))\right\}\right\}=\infty.$$

Because of this, we consider $\alpha>1$ in this section, .

$\underline{\delta>0}$. Assume $X_1,\ldots,X_\tau$ coming from $f(x;\mu,\sigma,\alpha)$, $X_{\tau+1},\ldots,X_n$.. coming from $f(x;\mu+\delta,\sigma,\alpha)$.

Theorem 3.2.5. *For the following hypothesis testing problem:*

$$H_0:\tau\geq n.$$
$$H_1:0\leq\tau<n.$$

Denote $Y_i=(\alpha-1)(\log(\mu+\delta-X_i)-\log(\mu-X_i))-(\sigma^{-1}(\mu+\delta-X_i))^{\alpha}+(\sigma^{-1}(\mu-X_i))^{\alpha}$, $S_k=\sum_{i=1}^{k}Y_i$, *we reject if either there exists* j, *such that* $X_j\geq\mu$, *or if* $S_n-\min_{0\leq k<n}S_k\geq L$ *under the condition that* $x_j<\mu$ *for all* j.

Proof. The likelihood for the null hypothesis is

$$L_0(\tau)=\prod_{i=1}^{n}\sigma^{-1}\alpha(-\sigma^{-1}(X_i-\mu))^{\alpha-1}\exp\left\{-(-\sigma^{-1}(X_i-\mu))^{\alpha}\right\}I_{(-\infty,\mu)}(X_i)$$

The likelihood for the alternative hypothesis is

$$\begin{aligned}L_1(\tau)=&\prod_{i=1}^{\tau}\sigma^{-1}\alpha(-\sigma^{-1}(X_i-\mu))^{\alpha-1}\exp\left\{-(-\sigma^{-1}(X_i-\mu))^{\alpha}\right\}I_{(-\infty,\mu)}(X_i)\\&\times\prod_{i=\tau+1}^{n}\sigma^{-1}\alpha(-\sigma^{-1}(X_i-\mu-\delta))^{\alpha-1}\exp\left\{-(-\sigma^{-1}(X_i-\mu-\delta))^{\alpha}\right\}I_{(-\infty,\mu+\delta)}(X_i)\end{aligned}$$

If there exists j, $1\leq j\leq n$, such that $X_j\geq\mu$, the the likelihood for the null is zero, and we reject the null hypothesis. If for all j, $X_j < \mu$, then the likelihood ratio statistic is

$$
\begin{aligned}
\Lambda(\tau) &= \frac{\max_{0\le\tau<n} L_1(\tau)}{\max_{\tau\ge n} L_0(\tau)} \\
&\propto \max_{0\le\tau<n} L_1(\tau) \\
&\propto \max_{0\le\tau<n} \prod_{i=1}^{\tau} \sigma^{-1}\alpha(-\sigma^{-1}(X_i-\mu))^{\alpha-1}\exp\left\{-(-\sigma^{-1}(X_i-\mu))^{\alpha}\right\} I_{(-\infty,\mu)}(X_i) \\
&\times \prod_{i=\tau+1}^{n} \sigma^{-1}\alpha(-\sigma^{-1}(X_i-\mu-\delta))^{\alpha-1}\exp\left\{-(-\sigma^{-1}(X_i-\mu-\delta))^{\alpha}\right\} I_{(-\infty,\mu+\delta)}(X_i) \\
&\propto \max_{0\le\tau<n} \prod_{i=1}^{\tau} (\mu-X_i)^{\alpha-1}\exp\left\{-(\sigma^{-1}(\mu-X_i))^{\alpha}\right\} \\
&\times \prod_{i=\tau+1}^{n} (\mu+\delta-X_i)^{\alpha-1}\exp\left\{-(\sigma^{-1}(\mu+\delta-X_i))^{\alpha}\right\}
\end{aligned}
$$

This is equivalent to maximizing

$$
\begin{aligned}
& \sum_{i=1}^{\tau}\{(\alpha-1)\log(\mu-X_i)-(\sigma^{-1}(\mu-X_i))^{\alpha}\} \\
+ & \sum_{i=\tau+1}^{n}\left\{(\alpha-1)\log(\mu+\delta-X_i)-(\sigma^{-1}(\mu+\delta-X_i))^{\alpha}\right\} \\
= & \sum_{i=1}^{\tau}\left\{(\alpha-1)\log(\mu-X_i)-(\sigma^{-1}(\mu-X_i))^{\alpha}\right\} \\
+ & \sum_{i=1}^{n}\left\{(\alpha-1)\log(\mu+\delta-X_i)-(\sigma^{-1}(\mu+\delta-X_i))^{\alpha}\right\} \\
- & \sum_{i=1}^{\tau}\left\{(\alpha-1)\log(\mu+\delta-X_i)-(\sigma^{-1}(\mu+\delta-X_i))^{\alpha}\right\} \\
= & \sum_{i=1}^{n}\left\{(\alpha-1)\log(\mu+\delta-X_i)-(\sigma^{-1}(\mu+\delta-X_i))^{\alpha}\right\} \\
- & \sum_{i=1}^{\tau}\left\{(\alpha-1)(\log(\mu+\delta-X_i)-\log(\mu-X_i))-(\sigma^{-1}(\mu+\delta-X_i))^{\alpha}+(\sigma^{-1}(\mu-X_i))^{\alpha}\right\}
\end{aligned}
$$

If we denote $Y_i = (\alpha-1)(\log(\mu+\delta-X_i)-\log(\mu-X_i))-(\sigma^{-1}(\mu+\delta-X_i))^{\alpha}+(\sigma^{-1}(\mu-X_i))^{\alpha}$, $S_k = \sum_{i=1}^{k} Y_i$, then we choose $\hat{\tau} = \arg\min_{0\le k<n} S_k$.

The likelihood ratio statistic is therefore

$$\begin{aligned}
\Lambda(\hat{\tau}) &= \frac{L_1(\hat{\tau})}{L_0(\hat{\tau})} \\
&= \left\{ \prod_{i=\hat{\tau}+1}^{n} \sigma^{-1}\alpha(-\sigma^{-1}(X_i-\mu))^{\alpha-1} \exp\left\{-(-\sigma^{-1}(X_i-\mu))^{\alpha}\right\} \right\}^{-1} \\
&\times \left\{ \prod_{i=\hat{\tau}+1}^{n} \sigma^{-1}\alpha(-\sigma^{-1}(X_i-\mu-\delta))^{\alpha-1} \exp\left\{-(-\sigma^{-1}(X_i-\mu-\delta))^{\alpha}\right\} \right\} \\
&= \exp\left\{ \sum_{i=\tau+1}^{n} Y_i \right\} \\
&= \exp\left\{ S_n - \min_{0\le k<n} S_k \right\}
\end{aligned}$$

To sum up, the likelihood ratio statistic is

$$\Lambda(\hat{\tau}) = \infty I_{\{\exists i, s.t. X_i \ge \mu\}} + (S_n - \min_{0\le k<n} S_k) I_{\{\forall i, X_i<\mu\}}.$$

and we reject if $\Lambda(\hat{\tau}) \ge L$ for some constant L. This means that we reject if either there exists j, such that $X_j \ge \mu$, or if $S_n - \min_{0\le k<n} S_k \ge L$ under the condition that $X_j < \mu$ for all j. □

$\underline{\delta < 0}$. Assume $X_1, \ldots, X_\tau$ coming from $\overset{\text{iid}}{\sim} f(x; \mu, \sigma, \alpha)$ and $X_{\tau+1}, \ldots, X_n$.. coming from $\overset{\text{iid}}{\sim}$ $f(x; \mu+\delta, \sigma, \alpha)$.

Theorem 3.2.6. *For the following hypothesis testing problem:*

$$H_0 : \tau \ge n.$$
$$H_1 : 0 \le \tau < n.$$

Define $k_0 = \inf\{t : X_{t+1}, \ldots, X_n < \mu+\delta\}$, $Y_i = (\sigma^{-1}(\mu - X_i))^{\alpha} - (\sigma^{-1}(\mu+\delta-X_i))^{\alpha} + (\alpha-1)\log(\mu+\delta-X_i) - (\alpha-1)\log(\mu-X_i)$, *and* $\tilde{S}_k = \sum_{i=k+1}^{n} Y_i$, *we reject if* $\max_{k_0\le k<n} \tilde{S}_k \ge L$.

Proof. The likelihood function for the null hypothesis is

$$L_0(\tau) = \prod_{i=1}^{n} \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(\mu-X_i))^{\alpha}\right\} (\sigma^{-1}(\mu-X_i))^{\alpha-1} I_{(-\infty,\mu)}(X_i)$$

The likelihood function for the alternative hypothesis is

$$\begin{aligned} L_1(\tau) &= \prod_{i=1}^{\tau} \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha}\right\}(\sigma^{-1}(\mu - X_i))^{\alpha-1} I_{(-\infty,\mu)}(X_i) \\ &\times \prod_{i=\tau+1}^{n} \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(\mu + \delta - X_i))^{\alpha}\right\}(\sigma^{-1}(\mu + \delta - X_i))^{\alpha-1} I_{(-\infty,\mu+\delta)}(X_i) \end{aligned}$$

The likelihood ratio statistic is

$$\begin{aligned} \Lambda(\tau) = \frac{\max_{0\le\tau<n} L_1(\tau)}{\max_{\tau\ge n} L_0(\tau)} &\propto \max_{0\le\tau<n} \prod_{i=1}^{\tau} \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha}\right\}(\sigma^{-1}(\mu - X_i))^{\alpha-1} I_{(-\infty,\mu)}(X_i) \\ &\times \prod_{i=\tau+1}^{n} \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(\mu + \delta - X_i))^{\alpha}\right\}(\sigma^{-1}(\mu + \delta - X_i))^{\alpha-1} I_{(-\infty,\mu+\delta)}(X_i) \end{aligned}$$

Maximizing $\lambda(\tau)$ is equivalent to maximizing L_1 over τ. Firstly,

$$\tau = \{t : X_{t+1}, \ldots, X_n < \mu + \delta\} \tag{3.2.11}$$

Otherwise, $L_1(\tau)$ would be zero. Let $k_0 = \inf\{t : X_{t+1}, \ldots, X_n < \mu + \delta\}$. Under (3.2.11), we wish to maximize

$$\prod_{i=1}^{\tau} \exp\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha}\right\}(\sigma^{-1}(\mu - X_i))^{\alpha-1} \prod_{i=\tau+1}^{n} \exp\left\{-(\sigma^{-1}(\mu + \delta - X_i))^{\alpha}\right\}(\sigma^{-1}(\mu + \delta - X_i))^{\alpha-1}$$

This is equivalent to maximizing

$$
\begin{aligned}
& \sum_{i=1}^{\tau}\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha} + (\alpha-1)\log(\sigma^{-1}(\mu - X_i))\right\} \\
+ & \sum_{i=\tau+1}^{n}\left\{-(\sigma^{-1}(\mu+\delta - X_i))^{\alpha} + (\alpha-1)\log(\sigma^{-1}(\mu+\delta - X_i))\right\} \\
= & \sum_{i=1}^{n}\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha} + (\alpha-1)\log(\sigma^{-1}(\mu - X_i))\right\} \\
+ & \sum_{i=\tau+1}^{n}\left\{-(\sigma^{-1}(\mu+\delta - X_i))^{\alpha} + (\alpha-1)\log(\sigma^{-1}(\mu+\delta - X_i))\right\} \\
+ & \sum_{i=\tau+1}^{n}\left\{(\sigma^{-1}(\mu - X_i))^{\alpha} - (\alpha-1)\log(\sigma^{-1}(\mu - X_i))\right\} \\
= & \sum_{i=1}^{n}\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha} + (\alpha-1)\log(\sigma^{-1}(\mu - X_i))\right\} \\
+ & \sum_{i=\tau+1}^{n}\left\{(\sigma^{-1}(\mu - X_i))^{\alpha} - (\sigma^{-1}(\mu+\delta - X_i))^{\alpha} + (\alpha-1)(\log(\mu+\delta - X_i) - \log(\mu - X_i))\right\}
\end{aligned}
$$

So if we let $Y_i = (\sigma^{-1}(\mu - X_i))^{\alpha} - (\sigma^{-1}(\mu+\delta - X_i))^{\alpha} + (\alpha-1)\log(\mu+\delta-X_i) - (\alpha-1)\log(\mu - X_i)$, we should choose $\hat{\tau}$ such that $\tilde{S}_k = \sum_{i=k+1}^{n} Y_i$ is maximized. From (3.2.11), we know that τ has to be at least k_0. Therefore we would choose $\hat{\tau} = \arg\max_{k_0 \le k < n} \tilde{S}_k$. Thus

$$
\begin{aligned}
\Lambda(\hat{\tau}) &= \exp\left\{\sum_{i=\hat{\tau}+1}^{n} Y_i\right\} \\
&= \exp\left\{\max_{k_0 \le k < n} \tilde{S}_k\right\}
\end{aligned}
$$

So we reject if $\max_{k_0 \le k < n} \tilde{S}_k \ge L$ for some constant L. $\square$

Change in α. Assume $X_1, \ldots, X_\tau$ comes from $f(x; \mu, \sigma, \alpha_1)$, $X_{\tau+1}, \ldots, X_n$.. comes from $f(x; \mu, \sigma, \alpha_2)$, where $\alpha_1, \alpha_2 > 1$.

Theorem 3.2.7. *For the following hypothesis testing problem:*

$$
\begin{aligned}
&H_0 : \tau \ge n. \\
&H_1 : 0 \le \tau < n.
\end{aligned}
$$

Define $Y_i = \log(\alpha_1^{-1}\alpha_2) - (\alpha_1 - \alpha_2)\log(\sigma^{-1}(\mu - X_i)) + (\sigma^{-1}(\mu - X_i))^{\alpha_1} - (\sigma^{-1}(\mu - X_i))^{\alpha_2}$, *and* $S_k = \sum_{i=1}^{k} Y_i$*, we reject if* $S_n - \min_{0\le k<n} S_k \ge L$*.*

Proof. The likelihood function for the null hypothesis is

$$L_0(\tau) = \prod_{i=1}^{n} \sigma^{-1}\alpha_1 \exp\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha_1}\right\}(\sigma^{-1}(\mu - X_i))^{\alpha_1 - 1} I_{(-\infty,\mu)}(X_i)$$

The likelihood function for the alternative hypothesis is

$$\begin{aligned} L_1(\tau) &= \prod_{i=1}^{\tau} \sigma^{-1}\alpha_1 \exp\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha_1}\right\}(\sigma^{-1}(\mu - X_i))^{\alpha_1 - 1} I_{(-\infty,\mu)}(X_i) \\ &\times \prod_{i=\tau+1}^{n} \sigma^{-1}\alpha_2 \exp\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha_2}\right\}(\sigma^{-1}(\mu - X_i))^{\alpha_2 - 1} I_{(-\infty,\mu)}(X_i) \end{aligned}$$

The likelihood ratio statistic is

$$\begin{aligned} \Lambda(\tau) = \frac{\max_{0\le\tau<n} L_1(\tau)}{\max_{\tau\ge n} L_0(\tau)} &\propto \max_{0\le\tau<n} \prod_{i=1}^{\tau} \sigma^{-1}\alpha_1 \exp\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha_1}\right\}(\sigma^{-1}(\mu - X_i))^{\alpha_1 - 1} I_{(-\infty,\mu)}(X_i) \\ &\times \prod_{i=\tau+1}^{n} \sigma^{-1}\alpha_2 \exp\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha_2}\right\}(\sigma^{-1}(\mu - X_i))^{\alpha_2 - 1} I_{(-\infty,\mu)}(X_i) \\ &\propto \max_{0\le\tau<n} \prod_{i=1}^{\tau} \alpha_1 \exp\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha_1}\right\}(\sigma^{-1}(\mu - X_i))^{\alpha_1} \\ &\times \prod_{i=\tau+1}^{n} \alpha_2 \exp\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha_2}\right\}(\sigma^{-1}(\mu - X_i))^{\alpha_2} \end{aligned}$$

This is equivalent to maximizing

$$\begin{aligned} & \sum_{i=1}^{\tau}\left\{\log\alpha_1 - (\sigma^{-1}(\mu - X_i))^{\alpha_1} + \alpha_1\log(\sigma^{-1}(\mu - X_i))\right\} \\ + & \sum_{i=\tau+1}^{n}\left\{\log\alpha_2 - (\sigma^{-1}(\mu - X_i))^{\alpha_2} + \alpha_2\log(\sigma^{-1}(\mu - X_i))\right\} \\ = & -\sum_{i=1}^{\tau}\left\{\log(\alpha_1^{-1}\alpha_2) - (\alpha_1 - \alpha_2)\log(\sigma^{-1}(\mu - X_i)) + (\sigma^{-1}(\mu - X_i))^{\alpha_1} - (\sigma^{-1}(\mu - X_i))^{\alpha_2}\right\} \\ + & \sum_{i=1}^{n}\left\{\log\alpha_2 - (\sigma^{-1}(\mu - X_i))^{\alpha_2} + \alpha_2\log(\sigma^{-1}(\mu - X_i))\right\} \end{aligned}$$

If we define $Y_i = \log(\alpha_1^{-1}\alpha_2) - (\alpha_1 - \alpha_2)\log(\sigma^{-1}(\mu - X_i)) + (\sigma^{-1}(\mu - X_i))^{\alpha_1} - (\sigma^{-1}(\mu - X_i))^{\alpha_2}$, and $S_k = \sum_{i=1}^{k} Y_i$, we choose $\hat{\tau} = \arg\min_{0\le k<n} S_k$.

$$\begin{aligned}
\Lambda(\hat{\tau}) &= \{\prod_{i=\hat{\tau}+1}^{n} \sigma^{-1}\alpha_1 \exp\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha_1}\right\}(\sigma^{-1}(\mu - X_i))^{\alpha_1 - 1} I_{(-\infty,\mu)}(X_i)\}^{-1} \\
&\times \prod_{i=\hat{\tau}+1}^{n} \sigma^{-1}\alpha_2 \exp\left\{-(\sigma^{-1}(\mu - X_i))^{\alpha_2}\right\}(\sigma^{-1}(\mu - X_i))^{\alpha_2 - 1} I_{(-\infty,\mu)}(X_i) \\
&= \exp\left\{\sum_{i=\hat{\tau}+1}^{n} Y_i\right\} = \exp\left\{S_n - \min_{0\le k<n} S_k\right\}
\end{aligned}$$

So the rejection region is $S_n - \min_{0\le k<n} S_k \ge L$ for some constant L. □

3.3 Generalized Pareto Distribution

The cumulative function for the generalized Pareto distribution is

$$\begin{cases}
1 - (1 + \sigma^{-1}\xi(x-\mu))^{-\frac{1}{\xi}} I_{[\mu,\infty)} & \xi > 0 \\
1 - (1 + \sigma^{-1}\xi(x-\mu))^{-\frac{1}{\xi}} I_{[\mu,\mu-\xi^{-1}\sigma]} & \xi < 0 \\
1 - \exp\left\{-\sigma^{-1}(x-\mu)\right\} & \xi = 0
\end{cases}$$

The probability density function is

$$\begin{cases}
\sigma^{-1}(1 + \sigma^{-1}\xi(x-\mu))^{-1-\frac{1}{\xi}} I_{[\mu,\infty)} & \xi > 0 \\
\sigma^{-1}(1 + \sigma^{-1}\xi(x-\mu))^{-1-\frac{1}{\xi}} I_{[\mu,\mu-\xi^{-1}\sigma]} & \xi < 0
\end{cases}$$

$\alpha > 0$, $\xi = 0$ and $\delta > 0$.

Assume $X_1, \ldots, X_\tau$ come from $\overset{\text{iid}}{\sim} f(x, \xi, \mu, \sigma_1) = \sigma_1^{-1}\exp\left\{-\sigma_1^{-1}(x-\mu)\right\} I_{[\mu,\infty)}(x)$, and $X_{\tau+1}, \ldots, X_n$.. come from $\overset{\text{iid}}{\sim} f(x, \xi, \mu+\delta, \sigma_2) = \sigma_2^{-1}\exp\left\{-\sigma_2^{-1}(x-\mu-\delta)\right\} I_{[\mu+\delta,\infty)}(x)$.

Theorem 3.3.1. *For the following hypothesis testing problem:*

$$H_0 : \tau \ge n.$$
$$H_1 : 0 \le \tau < n.$$

Define $k_0 = \inf\{t : X_{t+1}, X_{t+2}, \dots, X_n \geq \mu + \delta\}$, *and define* $Y_i = \sigma_1^{-1}(X_i - \mu) - \sigma_2^{-1}(X_i - \mu - \delta) + \log\sigma_1 - \log\sigma_2$, $\tilde{S}_k = \sum_{i=k+1}^{n} Y_i$, *We reject if* $\max_{k_0 \leq k < n} \tilde{S}_k \geq L$ *for some constant* L.

Proof. Let $k_0 = \inf\{t : X_{t+1}, X_{t+2}, \dots, X_n \geq \mu + \delta\}$. The likelihood for the null hypothesis is

$$L_0(\tau) = \prod_{i=1}^{n} \sigma_1^{-1} \exp\left\{-\sigma_1^{-1}(X_i - \mu)\right\} I_{[\mu,\infty)}(X_i),$$

The likelihood for the alternative hypothesis is

$$L_1(\tau) = \prod_{i=1}^{\tau} \sigma_1^{-1} \exp\left\{-\sigma_1^{-1}(X_i - \mu)\right\} I_{[\mu,\infty)}(X_i) \prod_{i=\tau+1}^{n} \sigma_2^{-1} \exp\left\{-\sigma_2^{-1}(X_i - \mu - \delta)\right\} I_{[\mu+\delta,\infty)}(X_i).$$

We see that τ must be at least k_0. So the likelihood ratio is

$$\begin{aligned}
\Lambda(\tau) &= \frac{\max_{0 \leq \tau < n} L_1(\tau)}{\max_{\tau \geq n} L_0(\tau)} \\
&\propto \max_{0 \leq \tau < n} \prod_{i=1}^{\tau} \sigma_1^{-1} \exp\left\{-\sigma_1^{-1}(X_i - \mu)\right\} I_{[\mu,\infty)}(X_i) \prod_{i=\tau+1}^{n} \sigma_2^{-1} \exp\left\{-\sigma_2^{-1}(X_i - \mu - \delta)\right\} I_{[\mu+\delta,\infty)}(X_i) \\
&\propto \max_{k_0 \leq \tau < n} \exp\left\{\sum_{i=1}^{\tau}[-\log\sigma_1 - \sigma_1^{-1}(X_i - \mu)] + \sum_{i=\tau+1}^{n}[-\log\sigma_2 - \sigma_2^{-1}(X_i - \mu - \delta)]\right\} \\
&= \max_{k_0 \leq \tau < n} \exp\{\sum_{i=\tau+1}^{n}\{\log\sigma_1 + \sigma_1^{-1}(X_i - \mu) - \log\sigma_2 - \sigma_2^{-1}(X_i - \mu - \delta)\} \\
&- \sum_{i=1}^{n}\{\log\sigma_1 + \sigma_1^{-1}(X_i - \mu)\}\}
\end{aligned}$$

To maximize $\Lambda(\tau)$, we need to maximize

$$\sum_{i=\tau+1}^{n}\{\log\sigma_1 + \sigma_1^{-1}(X_i - \mu) - \log\sigma_2 - \sigma_2^{-1}(X_i - \mu - \delta)\}$$

Define $Y_i = \log\sigma_1 + \sigma_1^{-1}(X_i - \mu) - \log\sigma_2 - \sigma_2^{-1}(X_i - \mu - \delta)$, and $\tilde{S}_k = \sum_{i=k+1}^{n} Y_i$, we choose $\hat{\tau} = \arg\max_{k_0 \leq k < n} \tilde{S}_k$.

In such case,

$$\begin{aligned}
\Lambda(\hat{\tau}) &= \left\{ \prod_{i=\hat{\tau}+1}^{n} \sigma_1^{-1} \exp\left\{-\sigma_1^{-1}(X_i-\mu)\right\} \right\}^{-1} \prod_{i=\hat{\tau}+1}^{n} \sigma_2^{-1} \exp\left\{-\sigma_2^{-1}(X_i-\mu-\delta)\right\} \\
&= \exp\left\{ \sum_{i=\hat{\tau}+1}^{n} Y_i \right\} \\
&= \exp\left\{ \max_{k_0 \le k < n} \tilde{S}_k \right\}
\end{aligned}$$

So the rejection region is $\Lambda(\hat{\tau}) \ge L$, which is $\max_{k_0 \le k < n} \tilde{S}_k \ge L$.

□

Corollary 3.3.2. *When $\sigma_1 = \sigma_2 = \sigma$, $Y_i = \sigma^{-1}\delta$ and $\tilde{S}_k = \sigma^{-1}(n-k)\delta$. Therefore $\max_{k_0 \le k < n} S_k = \sigma^{-1}(n-k_0)\delta$. The rejection criterion is expressed as $\sigma^{-1}\delta(n-k_0) \ge L$, which is essentially $n-k_0 \ge L$.*

<u>$\xi = 0$**,** $\delta < 0$ **and** $0 < \alpha < 1$</u>.

Assume that $X_1, \ldots, X_\tau$ come from $\stackrel{\text{iid}}{\sim} f(x,\xi,\mu,\sigma) = \sigma^{-1}\exp\left\{-\sigma^{-1}(x-\mu)\right\} I_{[\mu,\infty)}(x)$, and $X_{\tau+1}, \ldots, X_n, ..$ come from $f(x,\xi,\mu+\delta,\sigma) = \sigma^{-1}\exp\left\{-\sigma^{-1}(x-\mu-\delta)\right\} I_{[\mu+\delta,\infty)}(x)$.

Theorem 3.3.3. *For the following level α hypothesis testing problem:*

$$\begin{aligned}
H_0 &: \tau \ge n. \\
H_1 &: 0 \le \tau < n.
\end{aligned}$$

If there exists j, $1 \le j \le n$, such that $x_j < \mu$, H_1 holds; Otherwise H_0 holds.

Proof. The likelihood for the null hypothesis is

$$L_0(\tau) = \prod_{i=1}^{n} \sigma^{-1}\exp\left\{-\sigma^{-1}(X_i-\mu)\right\} I_{[\mu,\infty)}(X_i)$$

The likelihood for the alternative hypothesis is

$$L_1(\tau) = \prod_{i=1}^{\tau} \sigma^{-1}\exp\left\{-\sigma^{-1}(X_i-\mu)\right\} I_{[\mu,\infty)}(X_i) \prod_{i=\tau+1}^{n} \sigma^{-1}\exp\left\{-\sigma^{-1}(X_i-\mu-\delta)\right\} I_{[\mu+\delta,\infty)}(X_i)$$

If there exists j, $1 \le j \le n$, such that $X_j < \mu$, we reject the null hypothesis since the likelihood under the null will be zero.

If for all j, $X_j \geq \mu$, the likelihood ratio statistic becomes

$$\begin{aligned}
\Lambda(\tau) &= \frac{\max_{0\leq\tau<n} L_1(\tau)}{\max_{\tau\geq n} L_0(\tau)} \propto \max_{0\leq\tau<n} L_1(\tau) \\
&= \max_{0\leq\tau<n} \prod_{i=1}^{\tau} \sigma^{-1} \exp\left\{-\sigma^{-1}(X_i-\mu)\right\} \prod_{i=\tau+1}^{n} \sigma^{-1} \exp\left\{-\sigma^{-1}(X_i-\mu-\delta)\right\} \\
&= \max_{0\leq\tau<n} \exp\left\{-\sum_{i=1}^{\tau} \sigma^{-1}(X_i-\mu) - \sum_{i=\tau+1}^{n} \sigma^{-1}(X_i-\mu-\delta)\right\} \\
&= \max_{0\leq\tau<n} \exp\left\{-\sum_{i=1}^{n} \sigma^{-1}(X_i-\mu) + (n-\tau)\sigma^{-1}\delta\right\}
\end{aligned}$$

In order to maximize $\Lambda(\tau)$, we need to maximize τ because $\delta < 0$. Therefore $\hat{\tau} = n-1$.

$$\begin{aligned}
\Lambda(\hat{\tau}) &= \prod_{i=\hat{\tau}+1}^{n} \exp\left\{-\sigma^{-1}(X_i-\mu)\right\}^{-1} \exp\left\{-\sigma^{-1}(X_i-\mu-\delta)\right\} \\
&= \prod_{i=n}^{n} \exp\left\{\sigma^{-1}\delta\right\} = \exp\left\{\sigma^{-1}\delta\right\}
\end{aligned}$$

To sum up, $\Lambda(\hat{\tau}) = \infty I_{\{\exists j, s.t. X_j<\mu\}} + \exp\left\{\sigma^{-1}\delta\right\} I_{\{\forall j, s.t. X_j\geq\mu\}}$. We reject if $\Lambda(\hat{\tau}) \geq L$. If $L \leq \exp\left\{\sigma^{-1}\delta\right\}$, then $P_0(\exp\left\{\sigma^{-1}\delta\right\} \geq L) = 1 > \alpha$, which is contradictory to the level α test, therefore $L > \exp\left\{\sigma^{-1}\delta\right\}$. In this case, we reject only when there exists $1 \leq j \leq n$, such that $X_j < \mu$. $\square$

<u>$\xi_1 \neq \xi_2$**, and** $\xi_1, \xi_2 > 0$</u>.

Assume that $X_1, \ldots, X_\tau$ come from $\overset{\text{iid}}{\sim}$ from $f(x, \xi_1, \mu, \sigma) = \sigma^{-1}(1+\sigma^{-1}\xi_1(x-\mu))^{-1-\frac{1}{\xi_1}} I_{[\mu,\infty)}(x)$, and $X_{\tau+1}, \ldots, X_n$.. come from $\overset{\text{iid}}{\sim}$ from $f(x, \xi_2, \mu, \sigma) = \sigma^{-1}(1+\sigma^{-1}\xi_2(x-\mu))^{-1-\frac{1}{\xi_2}} I_{[\mu,\infty)}(x)$.

Theorem 3.3.4. *For the following hypothesis testing problem:*

$$\begin{aligned}
&H_0 : \tau \geq n. \\
&H_1 : 0 \leq \tau < n.
\end{aligned}$$

Define $Y_i = (1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) - (1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))$, *and* $S_k = \sum_{i=1}^{k} Y_i$, *we reject when* $S_n - \min_{0\leq k<n} S_k \geq L$ *for some constant* L.

Proof. The likelihood function for the null hypothesis is

$$L_0(\tau) = \prod_{i=1}^{n} \sigma^{-1}(1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}} I_{[\mu,\infty)}(X_i)$$

The likelihood function for the alternative hypothesis is

$$L_1(\tau) = \prod_{i=1}^{\tau} \sigma^{-1}(1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}} I_{[\mu,\infty)}(X_i) \prod_{i=\tau+1}^{n} \sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} I_{[\mu,\infty)}(X_i)$$

So the likelihood ratio statistic is

$$\begin{aligned}
\Lambda(\tau) &= \frac{\max_{0\le\tau<n} L_1(\tau)}{\max_{\tau\ge n} L_0(\tau)} \propto \max_{0\le\tau<n} L_1(\tau) \\
&= \max_{0\le\tau<n} \prod_{i=1}^{\tau} \sigma^{-1}(1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}} \prod_{i=\tau+1}^{n} \sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&\propto \max_{0\le\tau<n} \prod_{i=1}^{\tau} (1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}} \prod_{i=\tau+1}^{n} (1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}}
\end{aligned}$$

We want to maximize

$$\begin{aligned}
& \prod_{i=1}^{\tau}(1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}} \prod_{i=\tau+1}^{n}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
= \ & \exp\left\{\sum_{i=1}^{\tau}(-1-\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) + \sum_{i=\tau+1}^{n}(-1-\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))\right\} \\
= \ & \exp\{\sum_{i=1}^{\tau}(-1-\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) + \sum_{i=1}^{n}(-1-\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu)) \\
- \ & \sum_{i=1}^{\tau}(-1-\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))\} \\
= \ & \exp\{\sum_{i=1}^{n}(-1-\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu)) \\
- \ & \sum_{i=1}^{\tau}\{(1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) - (1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))\}\}
\end{aligned}$$

If we define $Y_i = (1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) - (1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))$, and $S_k = \sum_{i=1}^{k} Y_i$, we choose $\hat{\tau} = \arg\min_{0\le k<n} S_k$.
The likelihood statistic is therefore

$$\begin{aligned}
\Lambda(\hat{\tau}) &= \left\{\prod_{i=\tau+1}^{n}(1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}}\right\}^{-1} \prod_{i=\tau+1}^{n}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&= \exp\left\{\sum_{i=\tau+1}^{n} y_i\right\} = \exp\left\{S_n - \min_{0\le k<n} S_k\right\}
\end{aligned}$$

Therefore we reject if $S_n - \min_{0\le k<n} S_k \ge L$ for some constant L. □

$\xi_1 \neq \xi_2$, and $\xi_1 < \xi_2 < 0$.
Assume $X_1, \ldots, X_\tau$ coming from $f(x, \xi_1, \mu, \sigma) = \sigma^{-1}(1+\sigma^{-1}\xi_1(x-\mu))^{-1-\frac{1}{\xi_1}} I_{[\mu, \mu-\xi_1^{-1}\sigma]}(x)$, $X_{\tau+1}, \ldots, X_n$.. coming from $f(x, \xi_2, \mu, \sigma) = \sigma^{-1}(1+\sigma^{-1}\xi_2(x-\mu))^{-1-\frac{1}{\xi_2}} I_{[\mu, \mu-\xi_2^{-1}\sigma]}(x)$.

Theorem 3.3.5. *For the following hypothesis testing problem:*

$$H_0 : \tau \geq n.$$
$$H_1 : 0 \leq \tau < n.$$

Define $Y_i = (1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) - (1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))$, $S_k = \sum_{i=1}^k Y_i$. *If there exists* $1 \leq j \leq n$, $X_j > \mu - \xi_1^{-1}\sigma$, *we reject the null hypothesis. If for all* $1 \leq j \leq n$, $X_j \leq \mu - \xi_1^{-1}\sigma$, *then the rejection region is* $S_n - \min_{0 \leq k < n} S_k \geq L$ *for some constant* L.

Proof. The likelihood function for the null hypothesis is

$$L_0(\tau) = \prod_{i=1}^n \sigma^{-1}(1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}} I_{[\mu, \mu-\xi_1^{-1}\sigma]}(X_i)$$

The likelihood function for the alternative hypothesis is

$$L_1(\tau) = \prod_{i=1}^\tau \sigma^{-1}(1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}} I_{[\mu, \mu-\xi_1^{-1}\sigma]}(x_i) \prod_{i=\tau+1}^n \sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} I_{[\mu, \mu-\xi_2^{-1}\sigma]}(X_i)$$

Since $\xi_1 < \xi_2 < 0$, we have $-\xi_1^{-1}\sigma < -\xi_2^{-1}\sigma$. Therefore if there exists $1 \leq j \leq n$, $X_j > \mu - \xi_1^{-1}\sigma$, we reject the null hypothesis. If for all $1 \leq j \leq n$, $X_j \leq \mu - \xi_1^{-1}\sigma$, then the likelihood ratio statistic is

$$\begin{aligned}
\Lambda(\tau) &= \frac{\max_{0 \leq \tau < n} L_1(\tau)}{\max_{\tau \geq n} L_0(\tau)} \propto \max_{0 \leq \tau < n} L_1(\tau) \\
&= \max_{0 \leq \tau < n} \prod_{i=1}^\tau (1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}} \prod_{i=\tau+1}^n (1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&= \max_{0 \leq \tau < n} \exp\{\sum_{i=1}^\tau (-1-\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) \\
&+ \sum_{i=\tau+1}^n (-1-\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))\}
\end{aligned}$$

So we wish to minimize

$$\begin{aligned}
& \sum_{i=1}^{\tau}(1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) + \sum_{i=\tau+1}^{n}(1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu)) \\
= & \sum_{i=1}^{\tau}\left\{(1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) - (1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))\right\} \\
+ & \sum_{i=1}^{n}(1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))
\end{aligned}$$

If we define $Y_i = (1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) - (1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))$, $S_k = \sum_{i=1}^{k} Y_i$, we choose $\hat{\tau} = \arg\min_{0\le k<n} S_k$.

The likelihood ratio statistic is

$$\begin{aligned}
\Lambda(\hat{\tau}) &= \left\{\prod_{i=1}^{n}\sigma^{-1}(1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}} I_{[\mu,\mu-\xi_1^{-1}\sigma]}(X_i)\right\}^{-1} \\
&\times \prod_{i=1}^{\hat{\tau}}\sigma^{-1}(1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}} I_{[\mu,\mu-\xi_1^{-1}\sigma]}(X_i) \\
&\times \prod_{i=\hat{\tau}+1}^{n}\sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} I_{[\mu,\mu-\xi_2^{-1}\sigma]}(X_i) \\
&= \exp\left\{\sum_{i=\hat{\tau}+1}^{n} y_i\right\} = \exp\left\{S_n - \min_{0\le k<n} S_k\right\}
\end{aligned}$$

To sum up, the likelihood ratio statistic is

$$\Lambda(\hat{\tau}) = \infty I_{\{\exists 1\le j\le n, s.t. X_j>\mu-\xi_1^{-1}\sigma\}} + (S_n - \min_{0\le k<n} S_k) I_{\{\forall 1\le j\le n, X_j\le\mu-\xi_1^{-1}\sigma\}}$$

We reject if $\Lambda(\hat{\tau}) \ge L$ for some constant L. $\square$

$\xi_1 \neq \xi_2$, and $\xi_2 < \xi_1 < 0$. Assume that $X_1, \ldots, X_\tau$ come from $f(x,\xi_1,\mu,\sigma) = \sigma^{-1}(1+\sigma^{-1}\xi_1(x-\mu))^{-1-\frac{1}{\xi_1}} I_{[\mu,\mu-\xi_1^{-1}\sigma]}(x)$, and $X_{\tau+1}, \ldots, X_n$.. come from $f(x,\xi_2,\mu,\sigma) = \sigma^{-1}(1+\sigma^{-1}\xi_2(x-\mu))^{-1-\frac{1}{\xi_2}} I_{[\mu,\mu-\xi_2^{-1}\sigma]}(x)$.

Theorem 3.3.6. *For the following hypothesis testing problem:*

$$H_0 : \tau \ge n.$$
$$H_1 : 0 \le \tau < n.$$

Define $Y_i = (1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) - (1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))$, *and* $\tilde{S}_k =$

$\sum_{i=k+1}^{n} Y_i$, *and* $k_0 = \inf\left\{t : X_{t+1}, X_{t+2}, \ldots, x_n \leq \mu - \xi_2^{-1}\sigma\right\}$. *we reject if* $\max_{k_0 \leq k < n} \tilde{S}_k \geq L$.

Proof. The likelihood function under the null hypothesis is

$$L_0(\tau) = \prod_{i=1}^{n} \sigma^{-1}(1 + \sigma^{-1}\xi_1(X_i - \mu))^{-1-\frac{1}{\xi_1}} I_{[\mu, \mu - \xi_1^{-1}\sigma]}(X_i)$$

The likelihood function under the alternative hypothesis is

$$\begin{aligned} L_1(\tau) &= \prod_{i=1}^{\tau} \sigma^{-1}(1 + \sigma^{-1}\xi_1(X_i - \mu))^{-1-\frac{1}{\xi_1}} I_{[\mu, \mu - \xi_1^{-1}\sigma]}(X_i) \\ &\times \prod_{i=\tau+1}^{n} \sigma^{-1}(1 + \sigma^{-1}\xi_2(X_i - \mu))^{-1-\frac{1}{\xi_2}} I_{[\mu, \mu - \xi_2^{-1}\sigma]}(X_i) \end{aligned}$$

Since $\xi_2 < \xi_1 < 0$, so $\mu - \xi_1^{-1}\sigma > \mu - \xi_2^{-1}\sigma$. Let $k_0 = \inf\left\{t : X_{t+1}, X_{t+2}, \ldots, X_n \leq \mu - \xi_2^{-1}\sigma\right\}$. In this case, $\tau \geq k_0$. The likelihood ratio statistic is

$$\begin{aligned} \Lambda(\tau) &= \frac{\max_{0 \leq \tau < n} L_1(\tau)}{\max_{\tau \geq n} L_0(\tau)} \propto \max_{0 \leq \tau < n} L_1(\tau) \\ &= \max_{0 \leq \tau < n} \prod_{i=1}^{\tau} \sigma^{-1}(1 + \sigma^{-1}\xi_1(X_i - \mu))^{-1-\frac{1}{\xi_1}} \prod_{i=\tau+1}^{n} \sigma^{-1}(1 + \sigma^{-1}\xi_2(X_i - \mu))^{-1-\frac{1}{\xi_2}} \\ &\propto \max_{0 \leq \tau < n} \prod_{i=1}^{\tau} (1 + \sigma^{-1}\xi_1(X_i - \mu))^{-1-\frac{1}{\xi_1}} \prod_{i=\tau+1}^{n} (1 + \sigma^{-1}\xi_2(X_i - \mu))^{-1-\frac{1}{\xi_2}} \\ &= \max_{0 \leq \tau < n} \exp\{\sum_{i=1}^{\tau} (-1 - \xi_1^{-1}) \log(1 + \sigma^{-1}\xi_1(X_i - \mu)) \\ &+ \sum_{i=\tau+1}^{n} (-1 - \xi_2^{-1}) \log(1 + \sigma^{-1}\xi_2(X_i - \mu))\} \end{aligned}$$

So we are left to minimize

$$\begin{aligned}
& \sum_{i=1}^{\tau}(1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) + \sum_{i=\tau+1}^{n}(1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu)) \\
= & \sum_{i=1}^{n}(1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) + \sum_{i=\tau+1}^{n}(1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu)) \\
- & \sum_{i=\tau+1}^{n}(1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) \\
= & \sum_{i=1}^{n}(1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) \\
- & \sum_{i=\tau+1}^{n}\left\{(1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) - (1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))\right\}
\end{aligned}$$

Define $Y_i = (1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) - (1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))$, and $\tilde{S}_k = \sum_{i=k+1}^{n} Y_i$, we choose $\hat{\tau} = \arg\max_{k_0 \le k < n} \tilde{S}_k$. □

So the likelihood is

$$\begin{aligned}
\Lambda(\hat{\tau}) &= \left\{\prod_{i=1}^{n}\sigma^{-1}(1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}}\right\}^{-1}\prod_{i=1}^{\hat{\tau}}\sigma^{-1}(1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}} \\
&\times \prod_{i=\hat{\tau}+1}^{n}\sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&= \prod_{i=\tau+1}^{n}\left\{\sigma^{-1}(1+\sigma^{-1}\xi_1(X_i-\mu))^{-1-\frac{1}{\xi_1}}\right\}^{-1}\prod_{i=\hat{\tau}+1}^{n}\sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&= \exp\left\{\sum_{i=\hat{\tau}+1}^{n} y_i\right\} = \exp\left\{\max_{k_0\le k<n}\tilde{S}_k\right\}
\end{aligned}$$

We reject if $\max_{k_0\le k<n}\tilde{S}_k \ge L$ for some constant L.

$\underline{\xi_1 = 0, \xi_2 > 0}$.

Assume that $X_1,\ldots,X_\tau$ come from $\overset{\text{iid}}{\sim} f(x,\xi_1,\mu,\sigma) = \sigma^{-1}\exp\left\{-\sigma^{-1}(x-\mu)\right\} I_{[\mu,\infty)}(x)$, and $X_{\tau+1},\ldots,X_n$.. come from $\overset{\text{iid}}{\sim} f(x,\xi_2,\mu,\sigma) = \sigma^{-1}(1+\sigma^{-1}\xi_2(x-\mu))^{-1-\frac{1}{\xi_2}} I_{[\mu,\infty)}(x)$.

Theorem 3.3.7. *For the following hypothesis testing problem:*

$$H_0 : \tau \ge n.$$
$$H_1 : 0 \le \tau < n.$$

Define $Y_i = \sigma^{-1}(X_i - \mu) - (1 + \xi_2^{-1})\log(1 + \sigma^{-1}\xi_2(X_i - \mu))$, *and* $S_k = \sum_{i=1}^{k} Y_i$, *we reject when* $S_n - \min_{0\le k<n} S_k \ge L$ *for some constant* L.

Proof. The likelihood function for the null is

$$L_0(\tau) = \prod_{i=1}^{n} \sigma^{-1}\exp\left\{-\sigma^{-1}(X_i - \mu)\right\} I_{[\mu,\infty)}(X_i)$$

The likelihood function for the alternative is

$$L_1(\tau) = \prod_{i=1}^{\tau} \sigma^{-1}\exp\left\{-\sigma^{-1}(X_i - \mu)\right\} I_{[\mu,\infty)}(X_i) \prod_{i=\tau+1}^{n} \sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} I_{[\mu,\infty)}(X_i)$$

So the likelihood ratio statistic is

$$\begin{aligned}
\Lambda(\tau) &= \frac{\max_{0\le\tau<n} L_1(\tau)}{\max_{\tau\ge n} L_0(\tau)} \propto \max_{0\le\tau<n} L_1(\tau) \\
&= \prod_{i=1}^{\tau} \sigma^{-1}\exp\left\{-\sigma^{-1}(X_i - \mu)\right\} \prod_{i=\tau+1}^{n} \sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&\propto \prod_{i=1}^{\tau} \exp\left\{-\sigma^{-1}(X_i - \mu)\right\} \prod_{i=\tau+1}^{n} (1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&= \exp\left\{-\sum_{i=1}^{\tau}\sigma^{-1}(X_i-\mu) - \sum_{i=\tau+1}^{n}(1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))\right\} \\
&= \exp\{-\sum_{i=1}^{\tau}\sigma^{-1}(X_i-\mu) - \sum_{i=1}^{n}(1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu)) \\
&+ \sum_{i=1}^{\tau}(1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))\}
\end{aligned}$$

In order to maximize $\Lambda(\tau)$, we need to minimize $\sum_{i=1}^{\tau}\left\{\sigma^{-1}(X_i - \mu) - (1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))\right\}$. If we define $Y_i = \sigma^{-1}(X_i - \mu) - (1 + \xi_2^{-1})\log(1 + \sigma^{-1}\xi_2(X_i - \mu))$, and $S_k = \sum_{i=1}^{k} y_i$, we choose $\hat{\tau} = \arg\min_{0\le k<n} S_k$.

The likelihood statistic is therefore

$$\begin{aligned}
\Lambda(\hat{\tau}) &= \left\{\prod_{i=1}^{n} \sigma^{-1} \exp\left\{-\sigma^{-1}(X_i-\mu)\right\}\right\}^{-1} \\
&\times \prod_{i=1}^{\hat{\tau}} \sigma^{-1} \exp\left\{-\sigma^{-1}(X_i-\mu)\right\} \prod_{i=\hat{\tau}+1}^{n} \sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&= \left\{\prod_{i=\hat{\tau}+1}^{n} \sigma^{-1} \exp\left\{-\sigma^{-1}(X_i-\mu)\right\}\right\}^{-1} \prod_{i=\hat{\tau}+1}^{n} \sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&= \exp\left\{\sum_{i=\hat{\tau}+1}^{n} Y_i\right\} = \exp\left\{S_n - \min_{0\le k<n} S_k\right\}
\end{aligned}$$

we reject when $\Lambda(\hat{\tau}) \ge L$, which is equivalent to $S_n - \min_{0\le k<n} S_k \ge L$. □

$\underline{\xi_1 = 0, \xi_2 < 0}$.

Assume that $X_1, \ldots, X_\tau$ come from $\overset{\text{iid}}{\sim} f(x,\xi_1,\mu,\sigma) = \sigma^{-1}\exp\left\{-\sigma^{-1}(x-\mu)\right\} I_{[\mu,\infty)}(x)$, and $X_{\tau+1}, \ldots, X_n, ..$ come from $\overset{\text{iid}}{\sim} f(x,\xi_2,\mu,\sigma) = \sigma^{-1}(1+\sigma^{-1}\xi_2(x-\mu))^{-1-\frac{1}{\xi_2}} I_{[\mu,\mu-\xi_2^{-1}\sigma]}(x)$.

Theorem 3.3.8. *For the following hypothesis testing problem:*

$$H_0 : \tau \ge n.$$
$$H_1 : 0 \le \tau < n.$$

Define $k_0 = \inf\left\{t : X_{t+1}, X_{t+2}, \ldots, X_n \le \mu - \xi_2^{-1}\sigma\right\}$, $Y_i = \sigma^{-1}(X_i-\mu) - (1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))$, *and* $\tilde{S}_k = \sum_{i=k+1}^{n} y_i$, *we reject if* $\max_{k_0\le k<n} \tilde{S}_k \ge L$.

Proof. The likelihood function under the null hypothesis is

$$L_0(\tau) = \prod_{i=1}^{n} \sigma^{-1}\exp\{-\sigma^{-1}(X_i-\mu)\} I_{[\mu,\infty)}(X_i)$$

The likelihood function under the alternative is

$$L_1(\tau) = \prod_{i=1}^{\tau} \sigma^{-1}\exp\left\{-\sigma^{-1}(X_i-\mu)\right\} I_{[\mu,\infty)}(X_i) \prod_{i=\tau+1}^{n} \sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} I_{[\mu,\mu-\xi_2^{-1}\sigma]}(X_i)$$

define $k_0 = \inf\left\{t : X_{t+1}, X_{t+2}, \ldots, X_n \le \mu - \xi_2^{-1}\sigma\right\}$, so $\tau \ge k_0$.

The likelihood ratio statistic is therefore

$$\begin{aligned}
\Lambda(\tau) &= \frac{\max_{0\le\tau<n} L_1(\tau)}{\max_{\tau\ge n} L_0(\tau)} \propto \max_{0\le\tau<n} L_1(\tau) \\
&= \max_{k_0\le\tau<n} \prod_{i=1}^{\tau} \sigma^{-1}\exp\left\{-\sigma^{-1}(X_i-\mu)\right\} \prod_{i=\tau+1}^{n} \sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&\propto \max_{k_0\le\tau<n} \prod_{i=1}^{\tau} \exp\left\{-\sigma^{-1}(X_i-\mu)\right\} \prod_{i=\tau+1}^{n} (1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&= \max_{k_0\le\tau<n} \exp\left\{-\sum_{i=1}^{\tau}\sigma^{-1}(X_i-\mu) - \sum_{i=\tau+1}^{n}(1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))\right\}
\end{aligned}$$

So to maximize $\Lambda(\tau)$, we need to minimize

$$\begin{aligned}
&\sum_{i=1}^{\tau}\sigma^{-1}(X_i-\mu) + \sum_{i=\tau+1}^{n}(1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu)) \\
=&\sum_{i=1}^{n}\sigma^{-1}(X_i-\mu) - \sum_{i=\tau+1}^{n}\left\{\sigma^{-1}(X_i-\mu)-(1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))\right\}
\end{aligned}$$

If we define $Y_i = \sigma^{-1}(X_i-\mu)-(1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))$, and $\tilde{S}_k = \sum_{i=k+1}^{n} Y_i$, we choose $\hat{\tau} = \arg\max_{k_0\le k<n}\tilde{S}_k$.
The likelihood statistic is therefore

$$\begin{aligned}
\Lambda(\hat{\tau}) &= \left\{\prod_{i=1}^{n}\sigma^{-1}\exp\left\{-\sigma^{-1}(X_i-\mu)\right\}\right\}^{-1} \\
&\times \prod_{i=1}^{\hat{\tau}}\sigma^{-1}\exp\left\{-\sigma^{-1}(X_i-\mu)\right\}\prod_{i=\hat{\tau}+1}^{n}\sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&= \left\{\prod_{i=\hat{\tau}+1}^{n}\sigma^{-1}\exp\left\{-\sigma^{-1}(X_i-\mu)\right\}\right\}^{-1}\prod_{i=\hat{\tau}+1}^{n}\sigma^{-1}(1+\sigma^{-1}\xi_2(X_i-\mu))^{-1-\frac{1}{\xi_2}} \\
&= \exp\left\{\sum_{i=\hat{\tau}+1}^{n} Y_i\right\} = \exp\left\{\max_{k_0\le k<n}\tilde{S}_k\right\}
\end{aligned}$$

we reject when $\Lambda(\hat{\tau}) \ge L$, which is equivalent to $\max_{k_0\le k<n}\tilde{S}_k \ge L$. $\square$

Remark Theorem 3.3.6 and Theorem 3.3.8 has a strong connection in the sense that Theorem 3.3.8 is actually the limiting version of Theorem 3.3.6. Let $\xi_1 = -\epsilon$, where $\epsilon > 0$ is a small number. Based on Theorem 3.3.6,

$$Y_i^a = (1+\xi_1^{-1})\log(1+\sigma^{-1}\xi_1(X_i-\mu)) - (1+\xi_2^{-1})\log(1+\sigma^{-1}\xi_2(X_i-\mu))$$

Based on Theorem 3.3.8,

$$Y_i^b = \sigma^{-1}(X_i - \mu) - (1 + \xi_2^{-1}) \log(1 + \sigma^{-1}\xi_2(X_i - \mu))$$

We have

$$\begin{aligned} & Y_i^a - Y_i^b = (1 + \xi_1^{-1}) \log(1 + \sigma^{-1}\xi_1(X_i - \mu)) - \sigma^{-1}(X_i - \mu) \\ = \ & (1 - \epsilon^{-1}) \log(1 - \sigma^{-1}\epsilon(X_i - \mu)) - \sigma^{-1}(X_i - \mu) \\ = \ & (1 - \epsilon^{-1})(-\sigma^{-1}\epsilon(X_i - \mu) - o(\sigma^{-1}\epsilon(X_i - \mu)) - \sigma^{-1}(X_i - \mu) \\ = \ & -\sigma^{-1}\epsilon(X_i - \mu)) - o((\sigma^{-1}\epsilon(X_i - \mu))^2) \to 0 \end{aligned}$$

as $\epsilon \to 0$.
It is also the same case with Theorem 3.3.4 and Theorem 3.3.7, where we take $\xi_1 = \epsilon > 0$. By taking the limit of $\epsilon \to 0$, we get very similar results.

3.4 Summary: Change Detection on Maxima/Minima Statistic

This section summarizes the algorithm for change all of the parameters of interest: μ, σ and ξ. Assume that $X_1, X_2, \ldots, X_\tau \overset{\text{iid}}{\sim} f(x, \mu_1, \sigma_1, \xi_1)$, where

$$f(x, \mu_1, \sigma_1, \xi_1) = \frac{1}{\sigma_1}\{1 + \xi_1\frac{x_i - \mu_1}{\sigma_1}\}^{-1/\xi_1 - 1} \exp\{-\{1 + \xi_1\frac{x - \mu_1}{\sigma_1}\}^{-1/\xi_1}\} I_{\{1+\xi_1\frac{x-\mu_1}{\sigma_1}>0\}}$$

and $X_{\tau+1}, X_{\tau+2}, \ldots, X_n \overset{\text{iid}}{\sim} f(x, \mu_2, \sigma_2, \xi_2)$, where

$$f(x, \mu_2, \sigma_2, \xi_2) = \frac{1}{\sigma_2}\{1 + \xi_2\frac{x_i - \mu_2}{\sigma_2}\}^{-1/\xi_2 - 1} \exp\{-\{1 + \xi_2\frac{x - \mu_2}{\sigma_2}\}^{-1/\xi_2}\} I_{\{1+\xi_2\frac{x-\mu_2}{\sigma_2}>0\}}$$

We are interested in the hypothesis testing: $H_0 : \tau \geq n$ VS $H_1 : 0 \leq \tau < n$.

Since the density function differs according to the sign of ξ, we separate the discussion into four scenarios: a) $\xi_1 = 0, \xi_2 = 0$. b) $\xi_1 \neq 0, \xi_2 \neq 0$. c) $\xi_1 \neq 0, \xi_2 = 0$. d) $\xi_1 = 0, \xi_2 \neq 0$.

$\underline{\xi_1 = 0, \xi_2 = 0}$. It is included in Section 3.2.1 on Change in Gumbel distribution.

$\underline{\xi_1 \neq 0, \xi_2 \neq 0}$. The likelihood for the null hypothesis is:

$$\begin{aligned} & L_0(X_1, \ldots, X_n, \mu_1, \sigma_1, \xi_1) \\ = & \prod_{i=1}^{n} \frac{1}{\sigma_1}\{1 + \xi_1 \frac{X_i - \mu_1}{\sigma_1}\}^{-1/\xi_1 - 1} \exp\{-\{1 + \xi_1 \frac{X_i - \mu_1}{\sigma_1}\}^{-1/\xi_1}\} I_{\{1+\xi_1 \frac{X_i-\mu_1}{\sigma_1} > 0\}} \end{aligned}$$

The likelihood for the alternative is:

$$\begin{aligned} & L_1(X_1, \ldots, X_n, \mu_1, \sigma_1, \xi_1, \mu_2, \sigma_2, \xi_2) \\ = & \prod_{i=1}^{\tau} \frac{1}{\sigma_1}\{1 + \xi_1 \frac{X_i - \mu_1}{\sigma_1}\}^{-1/\xi_1 - 1} \exp\{-\{1 + \xi_1 \frac{X_i - \mu_1}{\sigma_1}\}^{-1/\xi_1}\} I_{\{1+\xi_1 \frac{X_i-\mu_1}{\sigma_1} > 0\}} \\ \times & \prod_{i=\tau+1}^{n} \frac{1}{\sigma_2}\{1 + \xi_2 \frac{X_i - \mu_2}{\sigma_2}\}^{-1/\xi_2 - 1} \exp\{-\{1 + \xi_2 \frac{X_i - \mu_2}{\sigma_2}\}^{-1/\xi_2}\} I_{\{1+\xi_2 \frac{X_i-\mu_2}{\sigma_2} > 0\}} \end{aligned}$$

If $\exists j$, such that $1 + \xi_1 \frac{X_j - \mu_1}{\sigma_1} \leq 0$, we reject the null hypothesis. Otherwise, $\forall j, 1 + \xi_1 \frac{X_j - \mu_1}{\sigma_1} > 0$, denote $k_0 = \inf\{t : 1 + \xi_2 \frac{X_{t+1} - \mu_2}{\sigma_2} > 0, 1 + \xi_2 \frac{X_{t+2} - \mu_2}{\sigma_2} > 0, \ldots, 1 + \xi_2 \frac{X_n - \mu_2}{\sigma_2} > 0\}$. Obviously, $\tau \geq k_0$.

The likelihood ratio statistic is

$$\begin{aligned} \Lambda(\tau) &= \frac{\max_{0 \leq \tau < n} L_1(\tau)}{\max_{\tau \geq n} L_0(\tau)} \propto \max_{0 \leq \tau < n} L_1(\tau) \\ &= \prod_{i=1}^{\tau} \frac{1}{\sigma_1}\{1 + \xi_1 \frac{X_i - \mu_1}{\sigma_1}\}^{-1/\xi_1 - 1} \exp\{-\{1 + \xi_1 \frac{X_i - \mu_1}{\sigma_1}\}^{-1/\xi_1}\} I_{\{1+\xi_1 \frac{X_i-\mu_1}{\sigma_1} > 0\}} \\ &\times \prod_{i=\tau+1}^{n} \frac{1}{\sigma_2}\{1 + \xi_2 \frac{X_i - \mu_2}{\sigma_2}\}^{-1/\xi_2 - 1} \exp\{-\{1 + \xi_2 \frac{X_i - \mu_2}{\sigma_2}\}^{-1/\xi_2}\} I_{\{1+\xi_2 \frac{X_i-\mu_2}{\sigma_2} > 0\}} \end{aligned}$$

Therefore, we need maximize

$$\begin{aligned}
& \sum_{i=1}^{\tau}\{-\log\sigma_1-(1+\frac{1}{\xi_1})\log(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})-(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})^{-1/\xi_1}\} \\
+ & \sum_{i=\tau+1}^{n}\{-\log\sigma_2-(1+\frac{1}{\xi_2})\log(1+\xi_2\frac{X_i-\mu_2}{\sigma_2})-(1+\xi_2\frac{X_i-\mu_2}{\sigma_2})^{-1/\xi_2}\} \\
= & \sum_{i=1}^{n}\{-\log\sigma_1-(1+\frac{1}{\xi_1})\log(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})-(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})^{-1/\xi_1}\} \\
- & \sum_{i=\tau+1}^{n}\{-\log\sigma_1-(1+\frac{1}{\xi_1})\log(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})-(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})^{-1/\xi_1}\} \\
+ & \sum_{i=\tau+1}^{n}\{-\log\sigma_2-(1+\frac{1}{\xi_2})\log(1+\xi_2\frac{X_i-\mu_2}{\sigma_2})-(1+\xi_2\frac{X_i-\mu_2}{\sigma_2})^{-1/\xi_2}\} \\
= & \sum_{i=1}^{n}\{-\log\sigma_1-(1+\frac{1}{\xi_1})\log(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})-(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})^{-1/\xi_1}\} \\
+ & \sum_{i=\tau+1}^{n}\{\log\sigma_1+(1+\frac{1}{\xi_1})\log(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})+(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})^{-1/\xi_1} \\
- & \log\sigma_2-(1+\frac{1}{\xi_2})\log(1+\xi_2\frac{X_i-\mu_2}{\sigma_2})-(1+\xi_2\frac{X_i-\mu_2}{\sigma_2})^{-1/\xi_2}\}
\end{aligned}$$

If we define $Y_i=\log\sigma_1+(1+\frac{1}{\xi_1})\log(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})+(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})^{-1/\xi_1}-\log\sigma_2-(1+\frac{1}{\xi_2})\log(1+\xi_2\frac{X_i-\mu_2}{\sigma_2})-(1+\xi_2\frac{X_i-\mu_2}{\sigma_2})^{-1/\xi_2}$, and $\tilde{S}_k=\sum_{i=k+1}^{n}Y_i$, then we need $\hat{\tau}=arg\max_{k_0\le k<n}\tilde{S}_k$.

The likelihood ratio becomes

$$\Lambda(\hat{\tau})=\exp\{\sum_{i=\hat{\tau}+1}^{n}Y_i\}=\exp\{\tilde{S}_{\hat{\tau}}\}=\exp\{\max_{k_0\le k<n}\tilde{S}_k\}$$

We conclude that If $\exists j$, such that $1+\xi_1\frac{X_j-\mu_1}{\sigma_1}\le 0$, we reject the null hypothesis. Otherwise, we reject if $\max_{k_0\le k<n}\tilde{S}_k\ge L$ for some constant L.

$\underline{\xi_1\neq 0,\xi_2=0}$. We use the result from b). Note that by taking $\xi\to 0$, we obtain Gumbel distribution. So if we take $\xi_2\to 0$ in the Y_i mentioned in b), we get $Y_i=\log\sigma_1+\frac{X_i-\mu_1}{\sigma_1}+\exp\{-\frac{X_i-\mu_1}{\sigma_1}\}-\log\sigma_2-(1+\frac{1}{\xi_2}\log(1+\xi_2\frac{X_i-\mu_2}{\sigma_2})-(1+\xi_2\frac{X_i-\mu_2}{\sigma_2})^{-1/\xi_2}$. The conclusion being if $\exists j$, such that $1+\xi_1\frac{X_j-\mu_1}{\sigma_1}\le 0$, we reject the null hypothesis. Otherwise, define $\tilde{S}_k=\sum_{i=k+1}^{n}Y_i$, we reject when $\max_{0\le k\le n}\tilde{S}_k\ge L$.

$\underline{\xi_1=0,\xi_2\neq 0}$. The approach is similar. Define $Y_i=\log\sigma_1+(1+\frac{1}{\xi_1})\log(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})+(1+\xi_1\frac{X_i-\mu_1}{\sigma_1})^{-1/\xi_1}-\log\sigma_2-\frac{X_i-\mu_2}{\sigma_2}-\exp\{-\frac{X_i-\mu_2}{\sigma_2}\}$. The conclusion is: define $k_0=\inf\{t:$

$1+\xi_2\frac{X_{t+1}-\mu_2}{\sigma_2} > 0, 1+\xi_2\frac{X_{t+2}-\mu_2}{\sigma_2} > 0, \ldots, 1+\xi_2\frac{X_n-\mu_2}{\sigma_2} > 0\}$, and $\tilde{S}_k = \sum_{i=k+1}^{n} y_i$, we reject if $\max_{k_0 \leq k \leq n} \tilde{S}_k \geq L$.

3.5 Simulation Study

This section provides a number of simulation studies on change of parameter(s) for the GEV distributions. Using both the Average Run Length approach and the p-value approach, we compare the performance of the GEV likelihood based procedure and the normality based procedure when the underlying distribution is a known GEV distribution. The only assumption we impose is that the first and the second moments of both distributions are equal.

3.5.1 Gumbel Simulation

Average Run Length approach. For a change from μ to $\mu+\delta$ in the Gumbel distribution under the average run length approach, the simulation procedures are summarized as follows. We simulate $B = 2500$ data series $\{X_n\}_{n=1}^{T} \overset{\text{iid}}{\sim} f(x;\mu,\sigma)$, where

$$f(x;\mu,\sigma) = \sigma^{-1}\exp\left\{-\sigma^{-1}(x-\mu)\right\}\exp\left\{-\exp\left\{-\sigma^{-1}(x-\mu)\right\}\right\}.$$

where $x \in R$ and $\sigma > 0$. Here T is fixed at 2000 for illustration as long as it is big enough. For the random variable X with Gumbel distribution $f(x;\mu,\sigma)$, we have

$$E(X) = \mu + \gamma\sigma, \qquad VAR(X) = \frac{1}{6}\sigma^2\pi^2. \tag{3.5.12}$$

Following Section 3.5.1, the Gumbel likelihood based decision rule is formulated as $S_n - \min_{0\leq k<n} S_k \geq L$. For a fixed L, and for each simulation $b = 1, \ldots, B = 2500$, we compute the value of R, where $R = inf\{n : S_n - min_{0\leq k<n} S_k \geq L\}$. $E_0(R)$ can be approximated if we compute the average of the R values based on these 2500 simulations. We search for L such that $\frac{|E_0(R)-200|}{200}$ is minimized. Since $E_0(R)$ is an increasing function of L, with values ranging from 0 to ∞, such an L exists, and it is unique.

Secondly, we compute $E_1((R-\tau)^+)$ under the alternative hypothesis. For each $\tau = 0, 1, \ldots, 100$, we simulate $X_1, \ldots, X_\tau \overset{\text{iid}}{\sim} f(x;\mu,\sigma)$ and $X_{\tau+1}, \ldots, X_T \overset{\text{iid}}{\sim} f(x;\mu+\delta,\sigma)$ for $B = 2500$ times, where δ is known. Here T is still kept as 2000.

Finally, we use the L in the first step and average the R values to get an approximation of $E_1((R-\tau)^+)$. Median and standard deviation of $(R-\tau)^+$ are also collected.

By Equation 3.5.12, the corresponding normality based scheme is formulated as a change from $N(\mu+\gamma\sigma, \frac{1}{6}\sigma^2\pi^2)$ to $N(\mu+\delta+\gamma\sigma, \frac{1}{6}\sigma^2\pi^2)$. We stop when we find the first n such that $T_n - \min_{0\le k<n} T_k \ge \tilde{L}$, where $T_k = \sum_{i=1}^{k} Z_i$, and $Z_i = -(X_i-\mu-\delta-\gamma\sigma)^2/(\frac{1}{3}\sigma^2\pi^2)+(X_i-\mu-\gamma\sigma)^2/(\frac{1}{3}\sigma^2\pi^2)$. We repeat exactly the same procedure as in the GEV case, computing $E_1((R-\tau)^+)$ for the normality based procedure and comparing it with the Gumbel based procedure.

For illustration, we set $\mu = 0$ and $\sigma = 1$. In this paper, four simulations are conducted with different δ, namely $\delta = 0.1$, $\delta = -0.1$, $\delta = 0.5$ and $\delta = -0.5$. See Figure 3.1 and 3.2. Studies show that when $\delta = 0.1$ or -0.1, for the Normality based scheme, it takes a very long time (close to 2000) to realize detection, indicating its inability to detect a small change. The Gumbel based procedure works very well, with an average run length of around 78. For a large change of $\delta = 0.5$ or -0.5, the advantage of the Gumbel based procedure is still obvious. For example, when $\delta = 0.5$, the average run length of the Gumbel based procedure is around 18, while the number is 27 for the normality counterpart.

P-value Approach. For a change from μ to $\mu+\delta$ in the Gumbel distribution under the p-value approach, the simulation procedure can be summarized as follows. As before, we simulate $\{X_n\}_{n=1}^{T} \overset{\text{iid}}{\sim} f(x;\mu,\sigma)$ for $B = 2500$ times, where T is fixed at 2000. Given each simulation $b = 1, 2, \ldots, B$, and for $n = 1, 2, \ldots, T$, we compute the Gumbel likelihood based statistic $S_n - \min_{0\le k<n} S_k$. Thus we have the empirical null distribution $\hat{F_n}$ formed by the 2500 values of $S_n - \min_{0\le k<n} S_k$.

Secondly, we compute $S_n - \min_{0\le k<n} S_k$ under the alternative distribution. Let τ be the change point. For illustration purposes, τ is set to be 20 and 120. We simulate $X_1, \ldots, X_\tau \overset{\text{iid}}{\sim} f(x;\mu,\sigma)$ and $X_{\tau+1}, \ldots, X_T \overset{\text{iid}}{\sim} f(x;\mu+\delta,\sigma)$ for 100 times, where δ is known. Here T is still kept as 2000. For any $\tau+1 \le n \le T$, we have 100 values of $S_n - \min_{0\le k<n} S_k$ under the alternative. For each of those values, a p-value can be calculated against the empirical null distribution $\hat{F_n}$. Thus, an empirical distribution of p-values across all $\tau+1 \le n \le T$ is obtained.

For the corresponding normality based scheme, we follow exactly the same procedure described above, except that we need to look at $T_n - \min_{0\le k<n} T_k$ under the null and the alternative, where T_k is derived from the normality change scheme in [14]. Empirical distribution plots of p-values for $\tau = 20$ and $\tau = 120$ are obtained as well.

As an example, assume $\mu = 0$ and $\sigma = 1$. We have performed four simulations with

$\delta = 0.1$, $\delta = -0.1$, $\delta = 0.5$ and $\delta = -0.5$. See Figure 3.8 through 3.15. These figures demonstrate the relative performance between the Gumbel likelihood based procedure and the normality based procedure in terms of 95%, 75%, 50% and 25% of the p-values.

From the simulation results, we notice that when the underlying distribution is Gumbel with parameters μ and σ, the p-value approach favors the Gumbel likelihood scheme rather than the normality based scheme. All of the 95%, 75%, 50% and 25% of the p-values exhibit the advantage of the Gumbel based scheme over the normality based scheme, and the advantage is dominant across different τ. It clearly indicates that the Gumbel based scheme is more effective at detecting a change than the normality based scheme. This phenomenon is most obvious when the change is subtle. Please compare the graph under $\delta = 0.1$ or -0.1 with the graph under $\delta = 0.5$ or -0.5.

3.5.2 Fréchet Simulation

Average Run Length Approach. For a change from μ to $\mu + \delta$ in the Fréchet distribution where $\delta > 0$, we simulate $B = 2500$ times of the data series $\{x_n\}_{n=1}^T$ coming from $f(x;\mu,\sigma,\alpha)$, where

$$f(x;\mu,\sigma,\alpha) = \sigma^{-1}\alpha \exp\left\{-(\sigma^{-1}(x-\mu))^{-\alpha}\right\}(\sigma^{-1}(x-\mu))^{-\alpha-1} I_{(\mu,\infty)}(x)$$

Here T is chosen as 2000. For illustration purposes, assume the distribution parameters are $\mu = 0, \sigma = 1$ and $\alpha = 3$. For such random variable X, both $E(X)$ and $VAR(X)$ exist, and

$$E(X) = \mu + \sigma\Gamma(1-1/\alpha) \qquad VAR(X) = \sigma^2\Gamma(1-2/\alpha) - \Gamma(1-1/\alpha)^2 \tag{3.5.13}$$

For a fixed L and each simulation $b = 1, 2, \ldots, B$, we simulate $\{x_j\}_{j=1}^T$. Following the discussion in Section 3.2.2, denote $R = \inf\{n : \max_{k_0 \le k < n} \tilde{S}_k \ge L\}$ with k_0 properly defined. In the case when all $1 \le n \le T$ satisfies $\max_{k_0 \le k < n} \tilde{S}_k < L$, we set $R = T$. For $B = 2500$ simulations, we obtain 2500 R values, and compute the average value of R, which serves as an approximation of $E_0(R)$. Then we search for L such that $\frac{|E_0(R)-200|}{200}$ is minimized.

Secondly, compute $E_1(R-\tau)^+$ under the alternative distribution. For $\tau = 1, 2, \ldots, 100$, simulate $X_1, \ldots, X_\tau \overset{\text{iid}}{\sim} f(x;\mu,\sigma,\alpha)$ and $X_{\tau+1}, \ldots, X_n \overset{\text{iid}}{\sim} f(x;\mu+\delta,\sigma,\alpha)$ for $B = 2500$ times, where δ is known. Here T is still kept as 2000. For each τ, use the L in the first step and compute R for $B = 2500$ times to obtain an approximation of $E_1(R-\tau)^+$. Median and standard deviation of $(R-\tau)^+$ are obtained in a similar way.

By Equation 3.5.13, the corresponding normality based scheme can be formulated as a change from $N(\mu+\sigma\Gamma(1-1/\alpha), \sigma^2(\Gamma(1-2/\alpha)-\Gamma(1-1/\alpha)^2))$ to $N(\mu+\delta+\sigma\Gamma(1-1/\alpha), \sigma^2(\Gamma(1-2/\alpha)-\Gamma(1-1/\alpha)^2))$. The detection rule is $T_n - \min_{0\leq k<n} T_k \geq \tilde{L}$, and T_k are defined as in [14]. We repeat exactly the same procedure for the normal case, computing $E_1((R-\tau)^+)$ and comparing it with the Fréchet based procedure.

We have completed simulations with $\delta = 0.1$ and $\delta = 0.5$. See Figure 3.3. Studies show that when $\delta = 0.1$, the average run length for the Fréchet based procedure is around 27, while the number is around 85 for the normality based procedure. When $\delta = 0.5$, the average run length for the Fréchet based procedure is around 5, while the number is around 23 for the normal counterpart. This indicates that the Fréchet based procedure statistic performs better than the normality based procedure, and the advantage is considerable when the change is small in size.

P-value Approach. We simulate $\{x_n\}_{n=1}^T \overset{\text{iid}}{\sim} f(x;\mu,\sigma,\alpha)$ for $B = 2500$ times. For each $n = 1, 2, \ldots, T$, if $X_n \leq \mu+\delta$, we assign $\max_{k_0\leq k<n} \tilde{S}_k = -\infty$. Otherwise, we compute the value for k_0 according to Section 3.2.2 as well as $\max_{k_0\leq k<n} \tilde{S}_k$. We repeat the process $B = 2500$ times. Therefore, for each n, we obtain $B = 2500$ values for $\max_{k_0\leq k<n} \tilde{S}_k$, which is the empirical distribution for $\max_{k_0\leq k<n} \tilde{S}_k$, denoted as $\hat{F}_n$.

Next, we simulate $X_1, \ldots, X_\tau \overset{\text{iid}}{\sim} f(x;\mu,\sigma,\alpha)$ and $X_{\tau+1}, \ldots, X_n \overset{\text{iid}}{\sim} f(x;\mu+\delta,\sigma,\alpha)$ for 100 times. Let $T = 2000$ and $\tau = 20$ and 120. For each $n = \tau+1, \ldots, T$, we assign the value of $\max_{k_0\leq k<n} \tilde{S}_k$. For the series $\{x_j\}_{j=1}^n$ where $n \geq \tau+1$, we compute k_0 and $\max_{k_0\leq k<n} \tilde{S}_k$. Therefore we get 100 values for $\max_{k_0\leq k<n} \tilde{S}_k$ under the alternative distribution. For each such value, a p-value can be calculated against the empirical null distribution $\hat{F}_n$. An empirical distribution plot of p-values is obtained across all $\tau+1 \leq n \leq T$.

For the corresponding normality based scheme, an empirical plot of the p-values for $\tau = 20$ and $\tau = 120$ is obtained as well. Figure 3.16 to 3.19 show the relative performance between the Fréchet based procedure and the normality based statistic in terms of the 95%, 75%, 50% and 25% of the p-values.

Average Run Length Approach. For a change from μ to $\mu+\delta$ in the Fréchet distribution where $\delta < 0$, we simulate $\{x_n\}_{n=1}^T$ coming from $f(x;\mu,\sigma,\alpha)$ for $B = 2500$ times, where T is fixed at 2000. For fixed L and each $n = 1, 2, \ldots, T$, R now becomes $\inf\{n : S_n - \min_{0\leq k<n} S_k \geq L \text{ or } x_n \leq \mu\}$. If for all $1 \leq n \leq T$, we observe $x_n > \mu$ and $S_n - \min_{0\leq k<n} S_k < L$, set $R = T$. The rest of the step follows exactly the same procedure as the $\delta > 0$ case for the Fréchet distribution change.

We conduct two simulations with $\delta = -0.1$ and $\delta = -0.5$. See Figure 3.4. Studies show that when $\delta = -0.5$, the average run length for the Fréchet likelihood based procedure is around 2, while the number is around 5 for the normality based procedure. When $\delta = -0.1$, the average run length for the Fréchet likelihood based procedure is around 20, while the number is around 49 for the normality based procedure.

P-value Approach. We simulate $\{x_n\}_{n=1}^{T} \overset{\text{iid}}{\sim} f(x; \mu, \sigma, \alpha)$ for $B = 2500$ times. For $n = 1, 2, \ldots, T$, we compute the $S_n - \min_{0 \le k < n} S_k$. For each $1 \le n \le T$, we obtain 2500 values for $S_n - \min_{0 \le k < n} S_k$, which constitutes the empirical null distribution $\hat{F}_n$.

Next, we simulate $\{x_n\}_{n=1}^{\tau} \overset{\text{iid}}{\sim} f(x; \mu, \sigma, \alpha)$, $\{x_n\}_{n=\tau+1}^{T} \overset{\text{iid}}{\sim} f(x; \mu + \delta, \sigma, \alpha)$ for 100 times. Let $T = 2000$ and $\tau = 20$ and 120. In each simulation, we assign the values of $S_n - \min_{0 \le k < n} S_k$ for $n = \tau + 1, \ldots, T$. For the series $\{x_j\}_{j=1}^{n}$ where $n \ge \tau + 1$, if $X_n \le \mu$, we assign $S_n - \min_{0 \le k < n} S_k = \infty$; otherwise we compute $S_n - \min_{0 \le k < n} S_k$. Therefore, we obtain 100 values for $S_n - \min_{0 \le k < n} S_k$ under the alternative distribution. For each such value, a p-value can be calculated against the empirical null distribution $\hat{F}_n$. Therefore, we get an empirical distribution plot of p-values across all $\tau + 1 \le n \le T$.

For the corresponding normality based scheme, an empirical plot of the p-values for $\tau = 20$ and $\tau = 120$ is obtained as well. Figure 3.20 to Figure 3.23 demonstrate the relative performance between the Fréchet based procedure and the normality based procedure in terms of the 95%, 75%, 50% and 25% of the p-values.

For a change from α to $\alpha+\delta$ in the Fréchet distribution, similar procedures are conducted, and the results for the average run length approach are summarized in Figure 3.5 and 3.6.

3.5.3 Generalized Pareto Simulation

The Generalized Pareto simulation study is very similar to the previous sections. In Figure 3.7, we consider a change from μ to $\mu + \delta$ for the Generalized Pareto distribution

$$f(x, \xi, \mu, \sigma_1) = \sigma_1^{-1} \exp\left\{-\sigma_1^{-1}(x - \mu)\right\} I_{[\mu,\infty)}(x)$$

where $\delta > 0$ and $\xi = 0$. Figure 3.24 to Figure 3.27 illustrate the p-value approach for $\tau = 20$ and 120, and show that it takes a much shorter time for the Generalized Pareto likelihood based procedure to signal an alarm than the normality based counterpart when a change occurs.

3.6 Hurricane Analysis with GEV

This section revisits the analysis on Atlantic hurricanes from year 1851 to 2008 by modeling the data as Generalized Extreme Value distribution. We focus on the changing behavior of the maximum sustained winds.

Let us consider the maximum sustained wind speeds as $\{x_n\}_{n=1}^{58}$. We treat it as coming from Gumbel $(\hat{\mu}, \hat{\sigma}_1)$ to Gumbel $(\hat{\mu}+\delta, \hat{\sigma}_2)$ where $\delta = c_1\hat{\sigma}_1$ and $\hat{\sigma}_2 = c_2\hat{\sigma}_1$. c_1 is predetermined as 0.25, 0.5 and 1; c_2 is predetermined as 1, 1.5 and 2. The Gumbel density function is therefore

$$f(y,\mu,\sigma) = \sigma^{-1}\exp\{-\sigma^{-1}(y_i-\mu)\}\exp\{-\exp\{-\sigma^{-1}(y_i-\mu)\}\} \tag{3.6.14}$$

The likelihood based methodology is introduced in Theorem 3.2.1. First, we need to estimate μ and σ_1 from the first 30 observations by the moment estimation method. Recalling Equation 3.5.12, we get

$$\begin{cases} \mu+\sigma_1\gamma = \dfrac{1}{30}\displaystyle\sum_{i=1}^{30} y_i \\ \dfrac{1}{6}\pi^2\sigma_1^2 = \dfrac{1}{29}\displaystyle\sum_{i=1}^{30}(y_i-\bar{Y})^2 \end{cases} \tag{3.6.15}$$

Therefore

$$\begin{cases} \hat{\sigma}_1 = \sqrt{\dfrac{6}{29\pi^2}\displaystyle\sum_{i=1}^{30}(y_i-\bar{Y})^2} = 11.01390 \\ \hat{\mu} = \dfrac{1}{30}\displaystyle\sum_{i=1}^{30} y_i - \hat{\sigma}_1\gamma = 96.30927 \end{cases} \tag{3.6.16}$$

Next, we simulate 2500 times $\{x_n\}_{n=1}^{T} \overset{\text{iid}}{\sim} f(y,\hat{\mu},\hat{\sigma}_1)$, where $f(y,\hat{\mu},\hat{\sigma}_1)$ is defined by Equation 3.6.14. Using the simulated data, we create the Gumbel likelihood ratio statistic $S_n - \min_{30\le k<n} S_k$ and obtain the L based on $E_0(R) = 200$. Here $t_i = \sigma_1^{-1}(y_i-\mu) - \sigma_2^{-1}(y_i-\mu-\delta) + \log\sigma_1 - \log\sigma_2 + \exp\{-\sigma_1^{-1}(y_i-\mu)\} - \exp\{-\sigma_2^{-1}(y_i-\mu)\}$ and $S_k = \sum_{i=1}^{k} t_i$ as defined in Theorem 3.2.1. Finally, we apply the hurricane data for the maximum sustained winds and search for the first n that makes $S_n - \min_{30\le k<n} S_k \ge L$. See table 3.1.

For the p-value approach on the maximum sustained wind speeds, we follow exactly the same procedure in Section 3.5.1 by computing the p-value of the Gumbel likelihood ratio statistic for the data against the simulated test statistics from our estimated parameters.

For details, please go to Table 3.2. The results are generally consistent, and both of the approaches indicate a change at the end of the 19th century to the beginning of the 20th century.

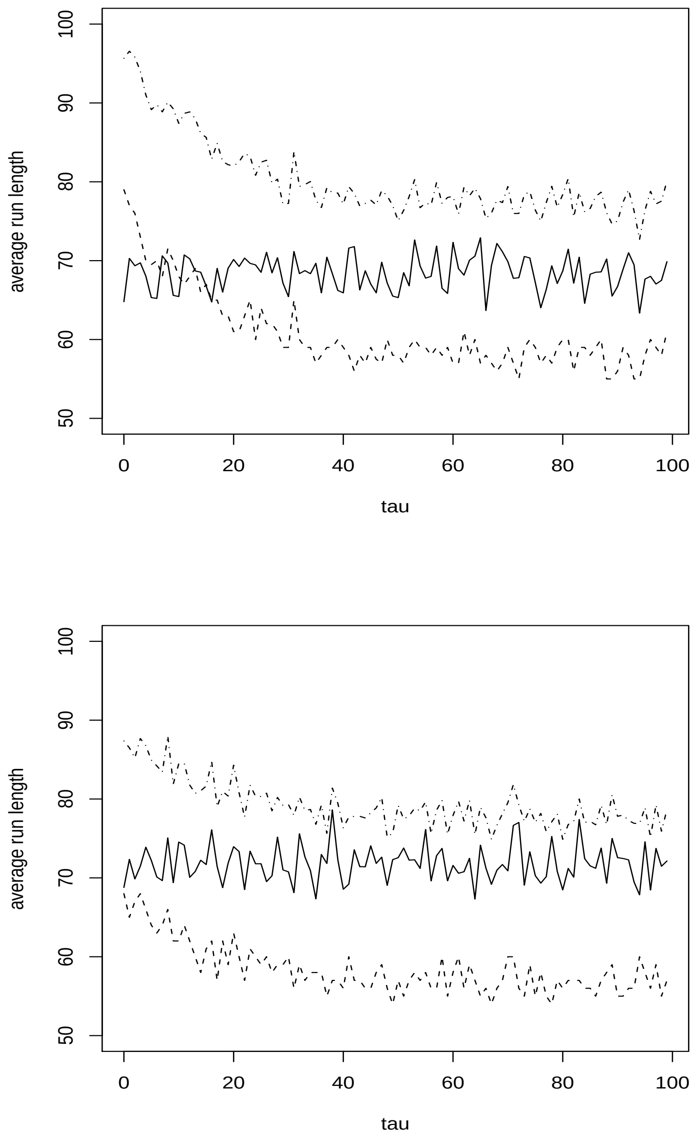


Figure 3.1: Gumbel likelihood ratio performance: The top panel shows change in Gumbel distribution with parameter change from $(\mu, \sigma) = (0, 1)$ to $(0.1, 1)$. The bottom panel shows change in Gumbel distribution with parameter change from $(\mu, \sigma) = (0, 1)$ to $(-0.1, 1)$. Dotdash, dashed and solid line stand for mean, median and standard deviation. The normal likelihood performance in both situations is not shown here because it gives an average run length of close to 2000, indicating its poor performance.

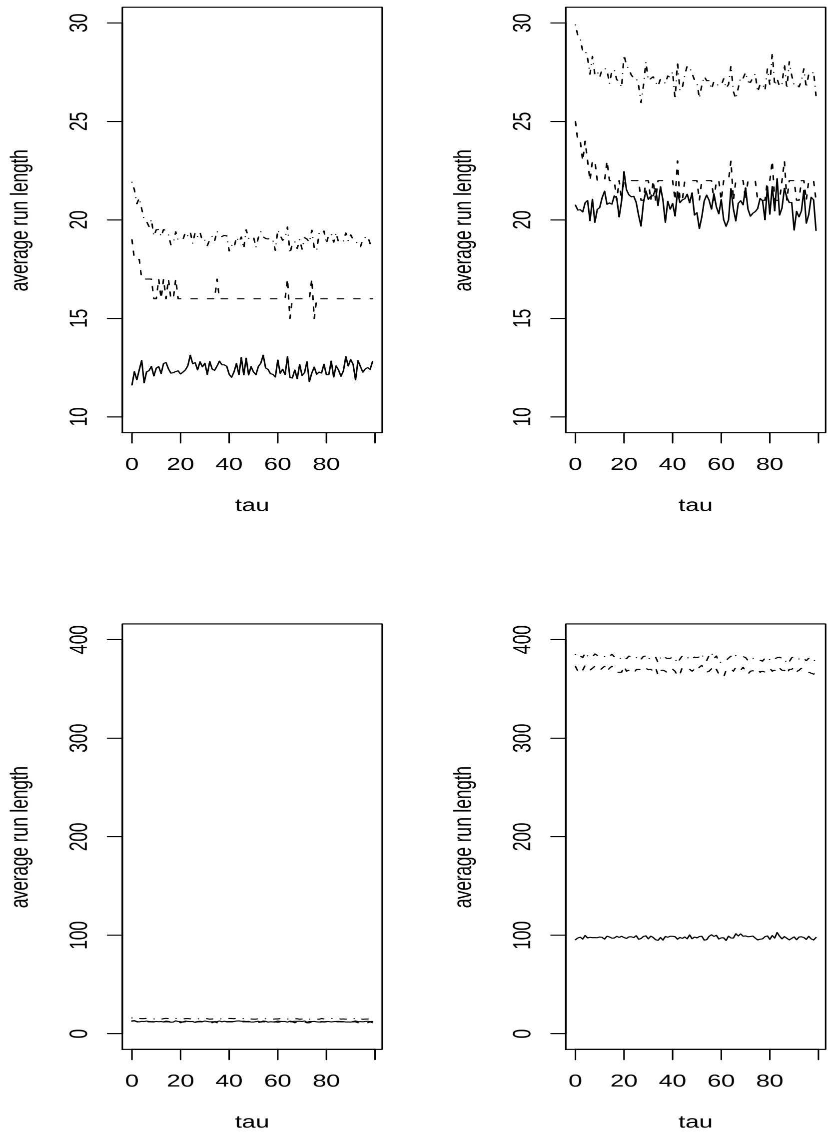


Figure 3.2: Performance Comparison: Gumbel likelihood ratio procedure with Normal likelihood ratio procedure. Dotdash, dashed and solid line stand for mean, median and standard deviation. The top panel describes change in Gumbel distribution with parameter change from $(\mu, \sigma) = (0, 1)$ to $(0.5, 1)$. The bottom panel describes change in Gumbel distribution with parameter change from $(\mu, \sigma) = (0, 1)$ to $(-0.5, 1)$.

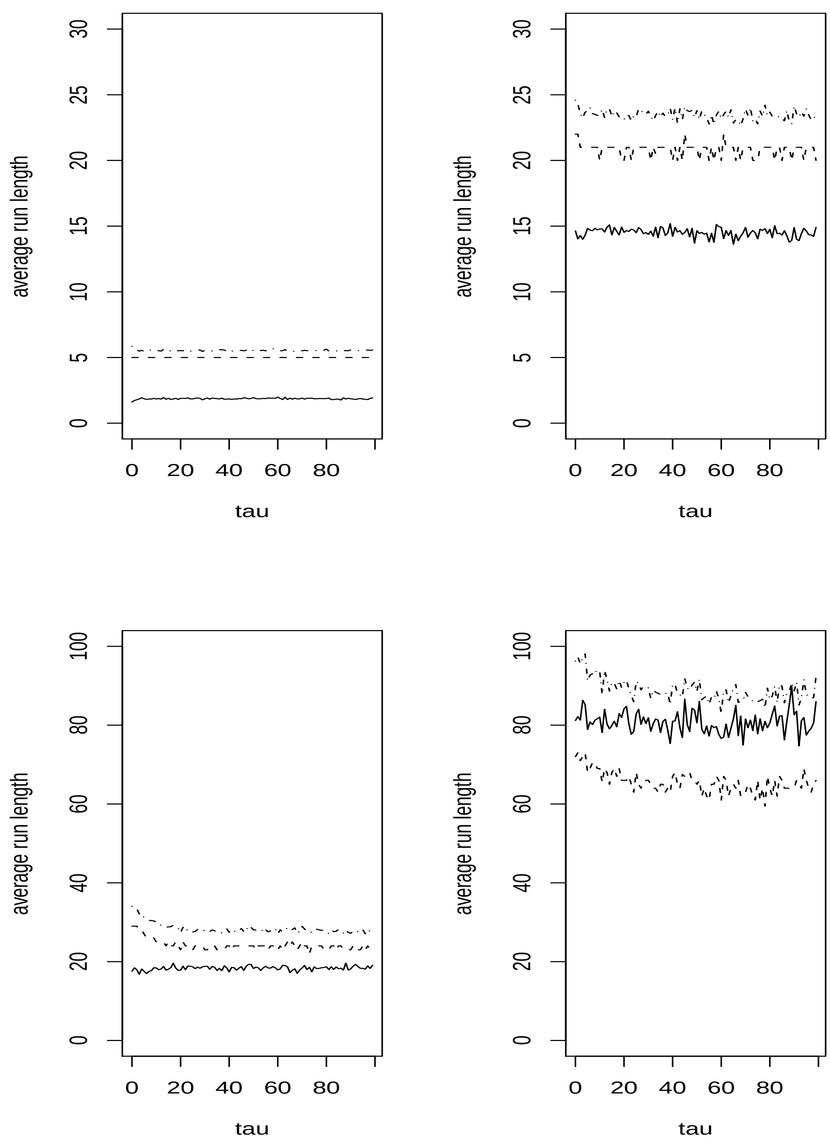


Figure 3.3: Performance Comparison: Fréchet likelihood ratio procedure with Normal likelihood ratio procedure. Dotdash, dashed and solid line stand for mean, median and standard deviation. The top panel describes change in Fréchet distribution with parameters change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(0.5, 1, 3)$. The bottom panel describes change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(0.1, 1, 3)$.

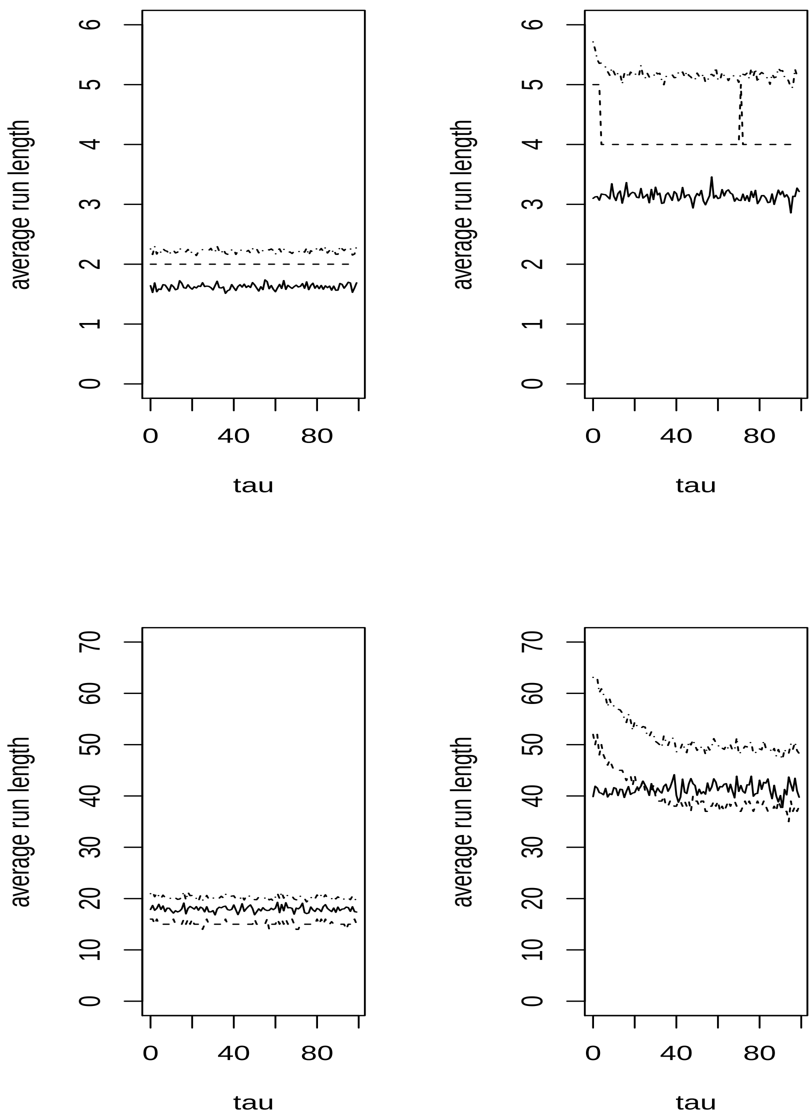


Figure 3.4: Performance Comparison: Fréchet likelihood ratio procedure with Normal likelihood ratio procedure. Dotdash, dashed and solid line stand for mean, median and standard deviation. The top panel describes change in Fréchet distribution with parameters change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(-0.5, 1, 3)$. The bottom panel describes change in Fréchet distribution with parameters change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(-0.1, 1, 3)$.

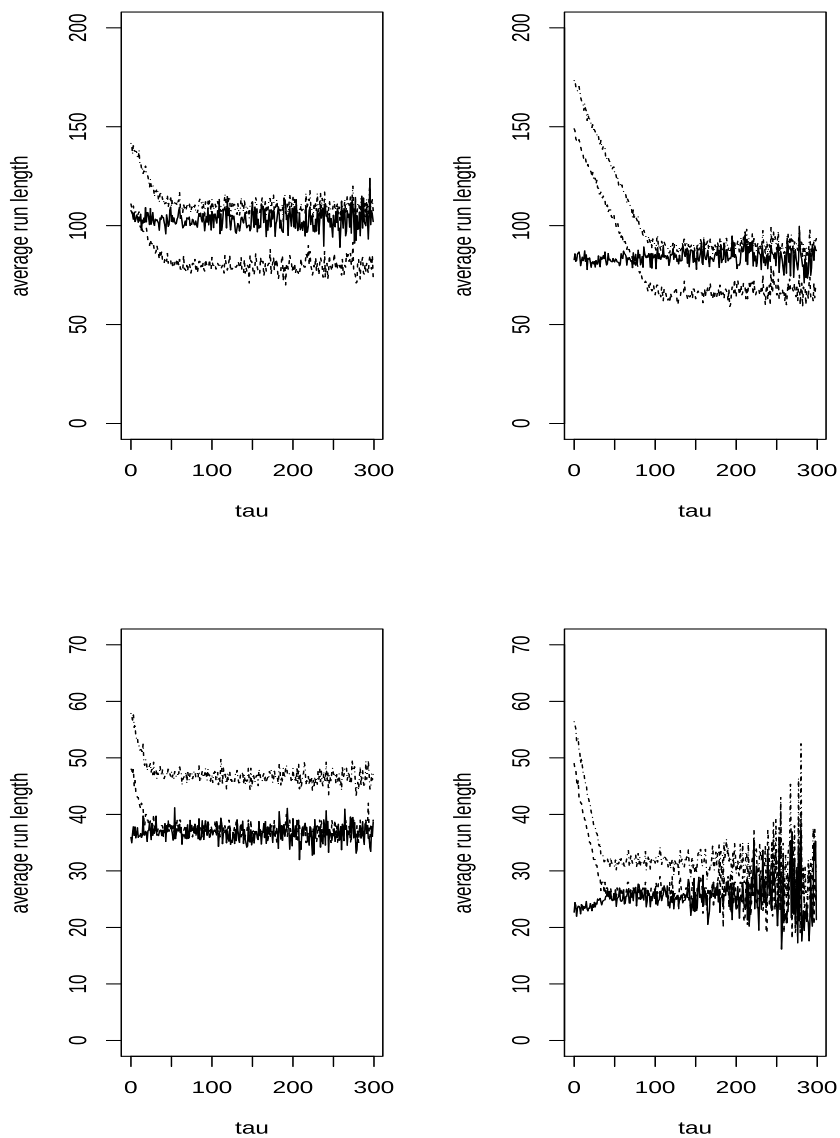


Figure 3.5: Performance Comparison: Fréchet likelihood ratio procedure with Normal likelihood ratio procedure. Dotdash, dashed and solid line stand for mean, median and standard deviation. The top panel describes change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3) \rightarrow (0, 1, 3.1)$, The bottom panel describes change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(0, 1, 3.5)$.

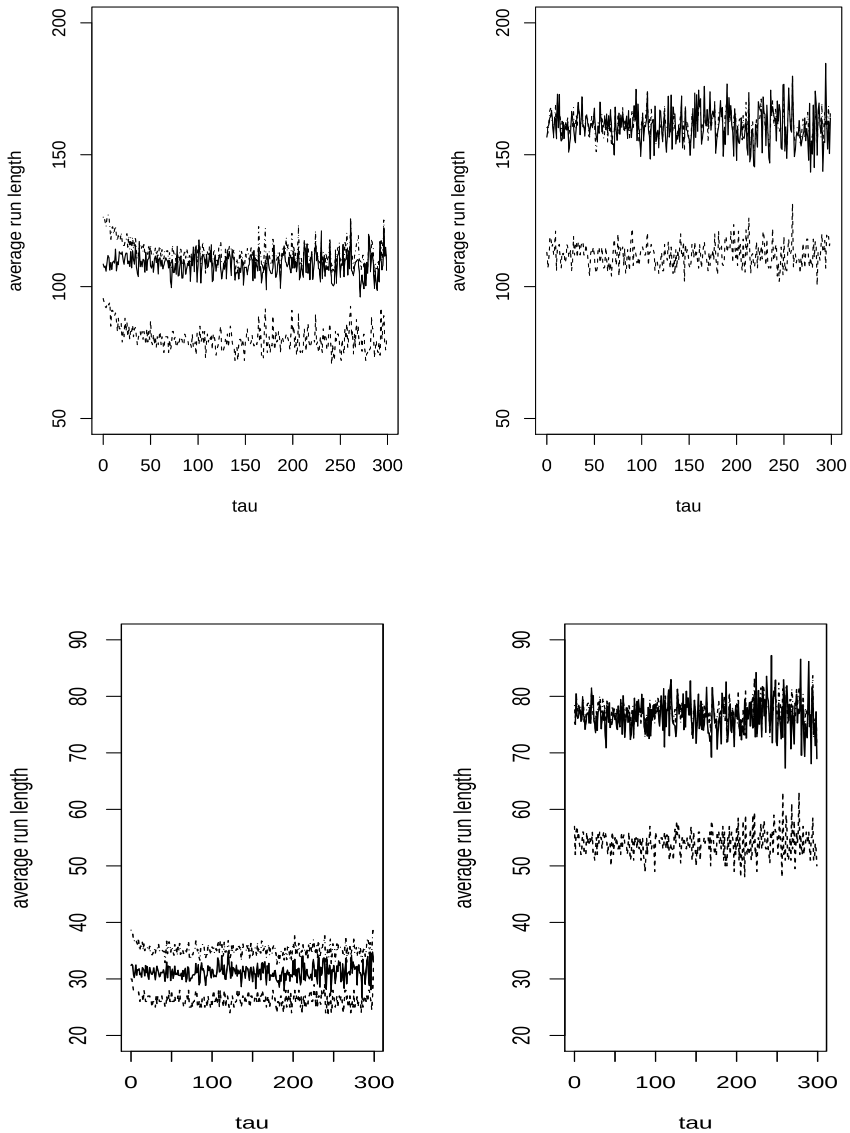


Figure 3.6: Performance Comparison: Fréchet likelihood ratio procedure with Normal likelihood ratio procedure. Dotdash, dashed and solid line stand for mean, median and standard deviation. The top panel describes change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(0, 1, 2.9)$, The bottom panel describes change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(0, 1, 2.5)$.

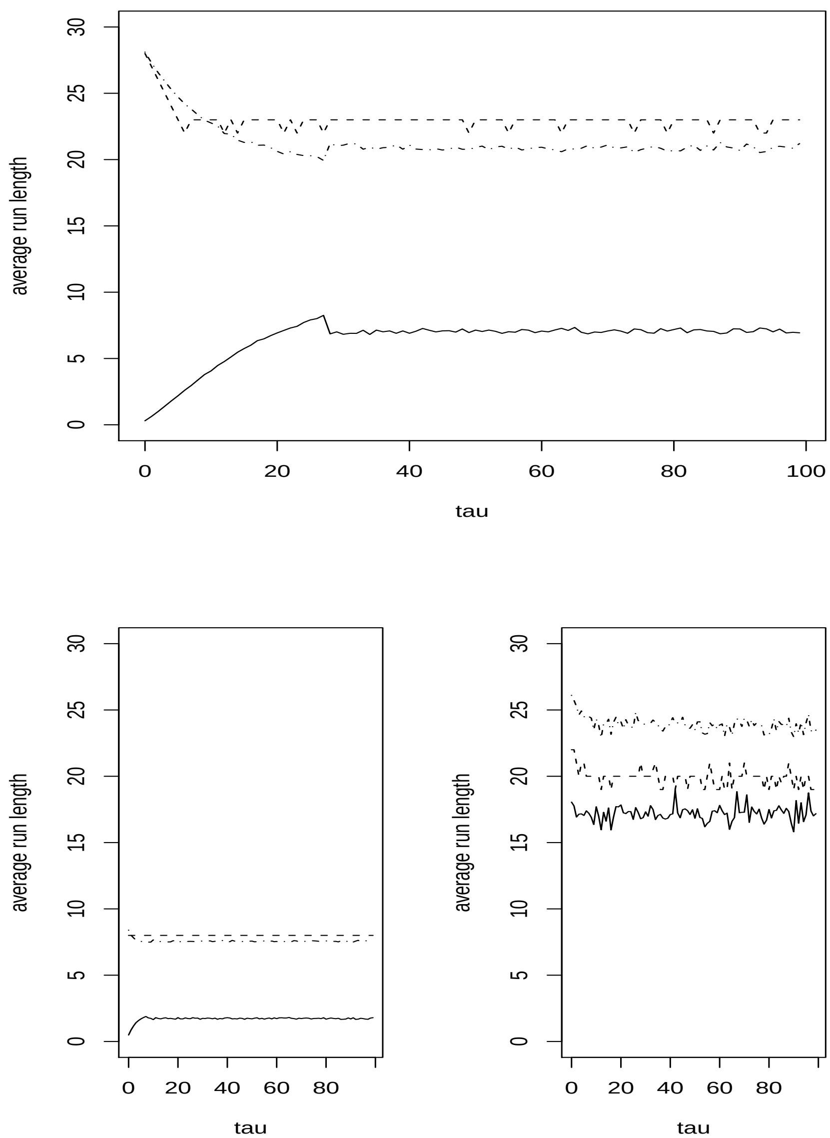


Figure 3.7: Generalized Pareto likelihood ratio procedure and Normal likelihood ratio procedure comparison. Dotdash, dashed and solid line stand for mean, median and standard deviation. The top panel shows change in generalized Pareto distribution with parameter change from $(\mu, \sigma) = (0, 1)$ to $(0.1, 1)$. The normal likelihood ratio procedure performance is not shown here because it gives an average run length of close to 2000, indicating its poor performance. The bottom panel shows change in generalized Pareto distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1)$ to $(0.5, 1)$.

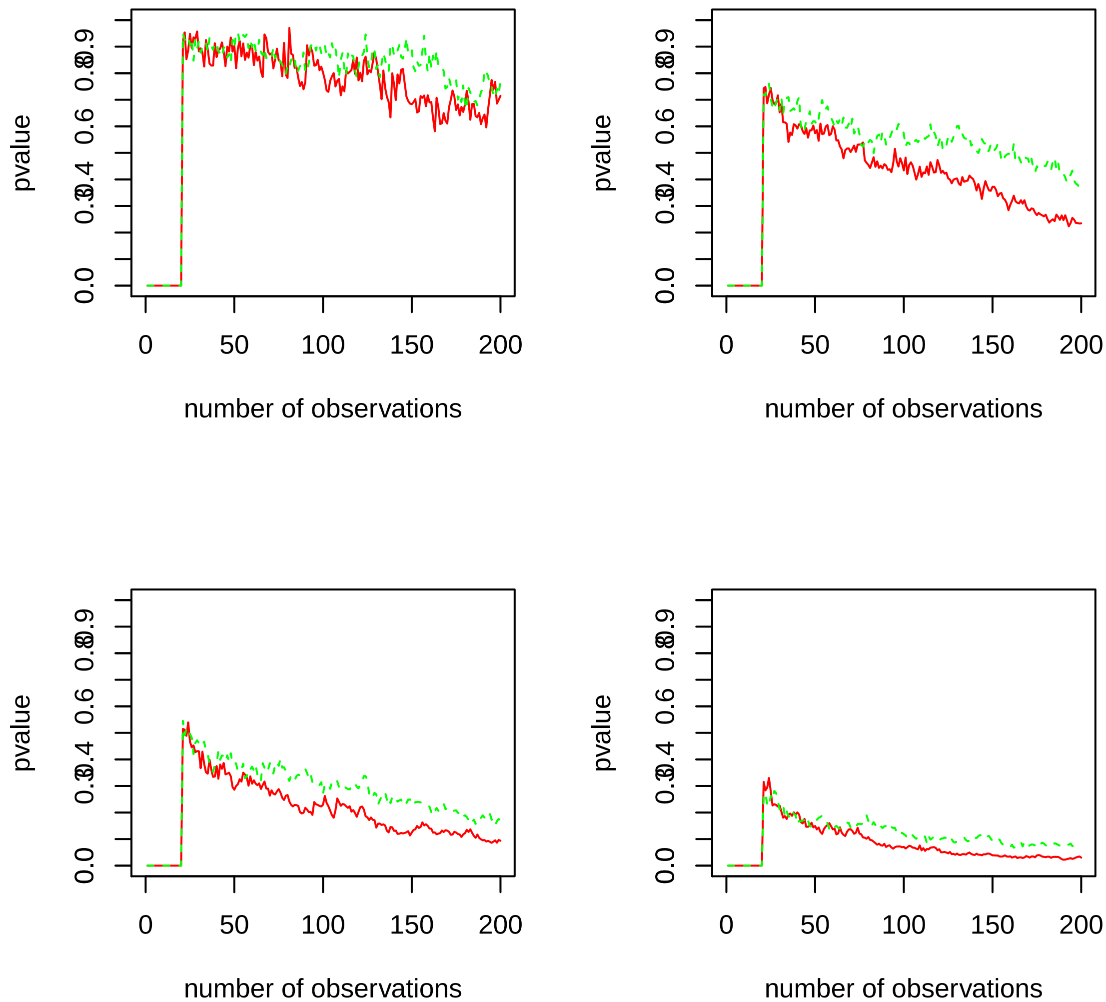


Figure 3.8: P-value Comparison. $\tau = 20$. Change in Gumbel distribution with parameter change from $(\mu, \sigma) = (0, 1)$ to $(0.1, 1)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Gumbel likelihood ratio procedure versus normal likelihood ratio procedure. The Gumbel likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

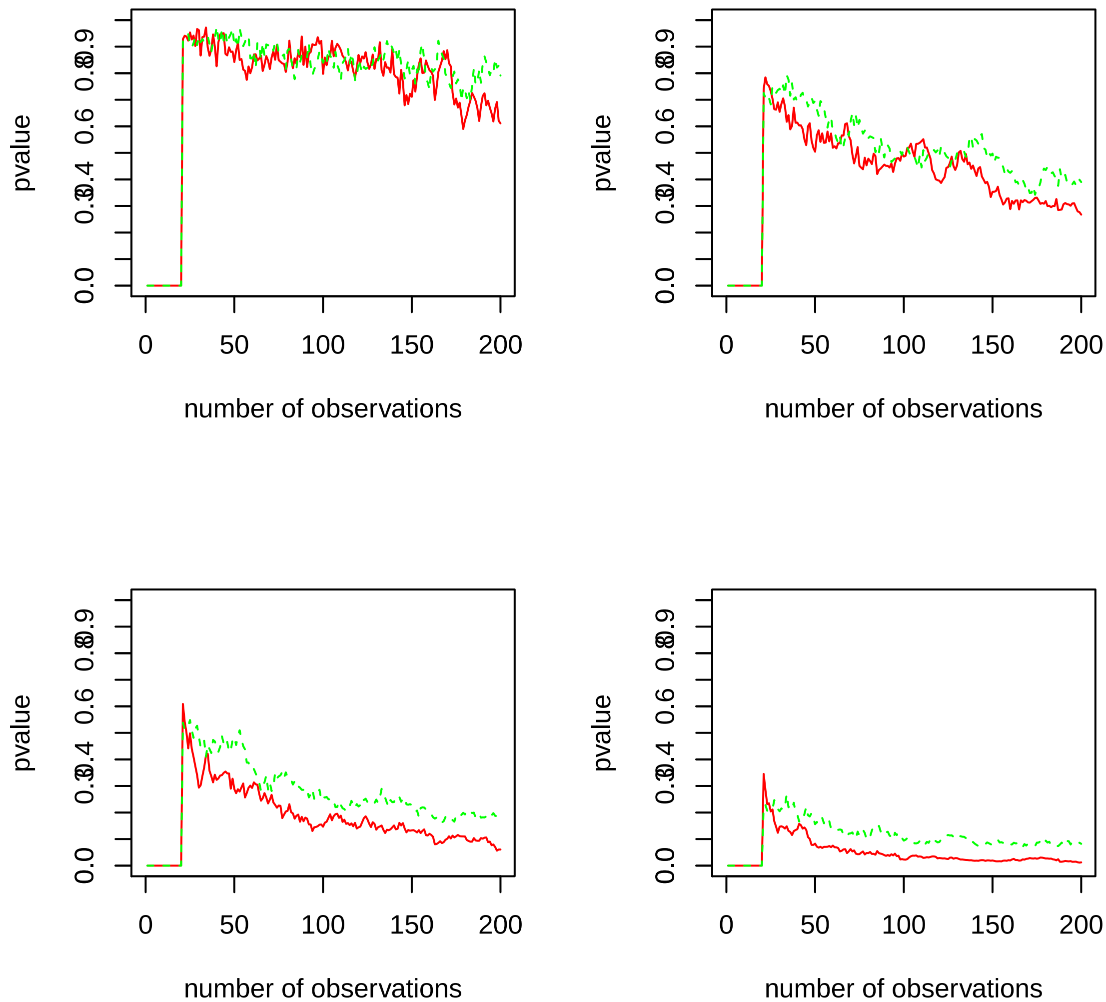


Figure 3.9: P-value Comparison. $\tau = 20$. Change in Gumbel distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1)$ to $(-0.1, 1)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Gumbel likelihood ratio procedure and normal likelihood ratio procedure. The Gumbel likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

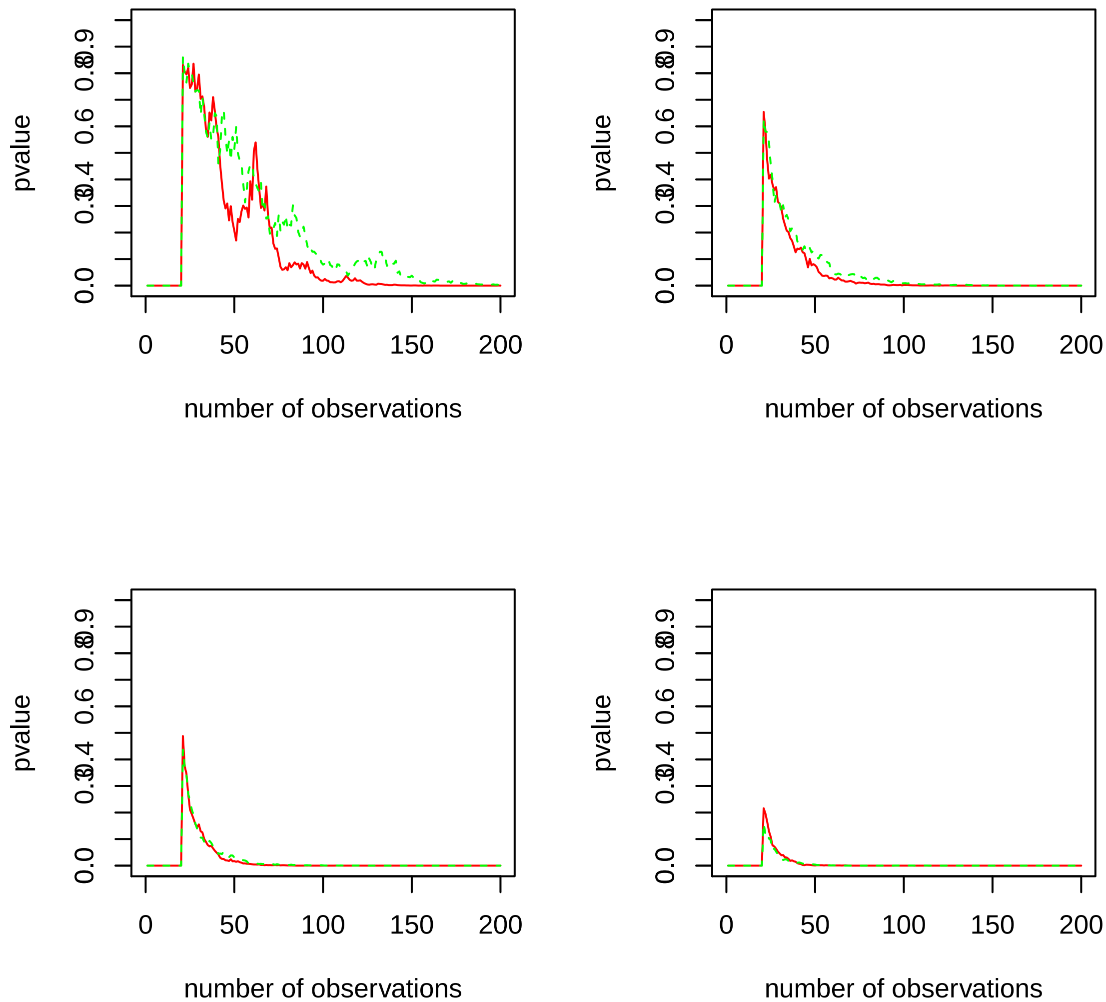


Figure 3.10: P-value Comparison. $\tau = 20$. Change in Gumbel distribution with parameter change from $(\mu, \sigma) = (0, 1)$ to $(0.5, 1)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Gumbel likelihood ratio procedure and normal likelihood ratio procedure. The Gumbel likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

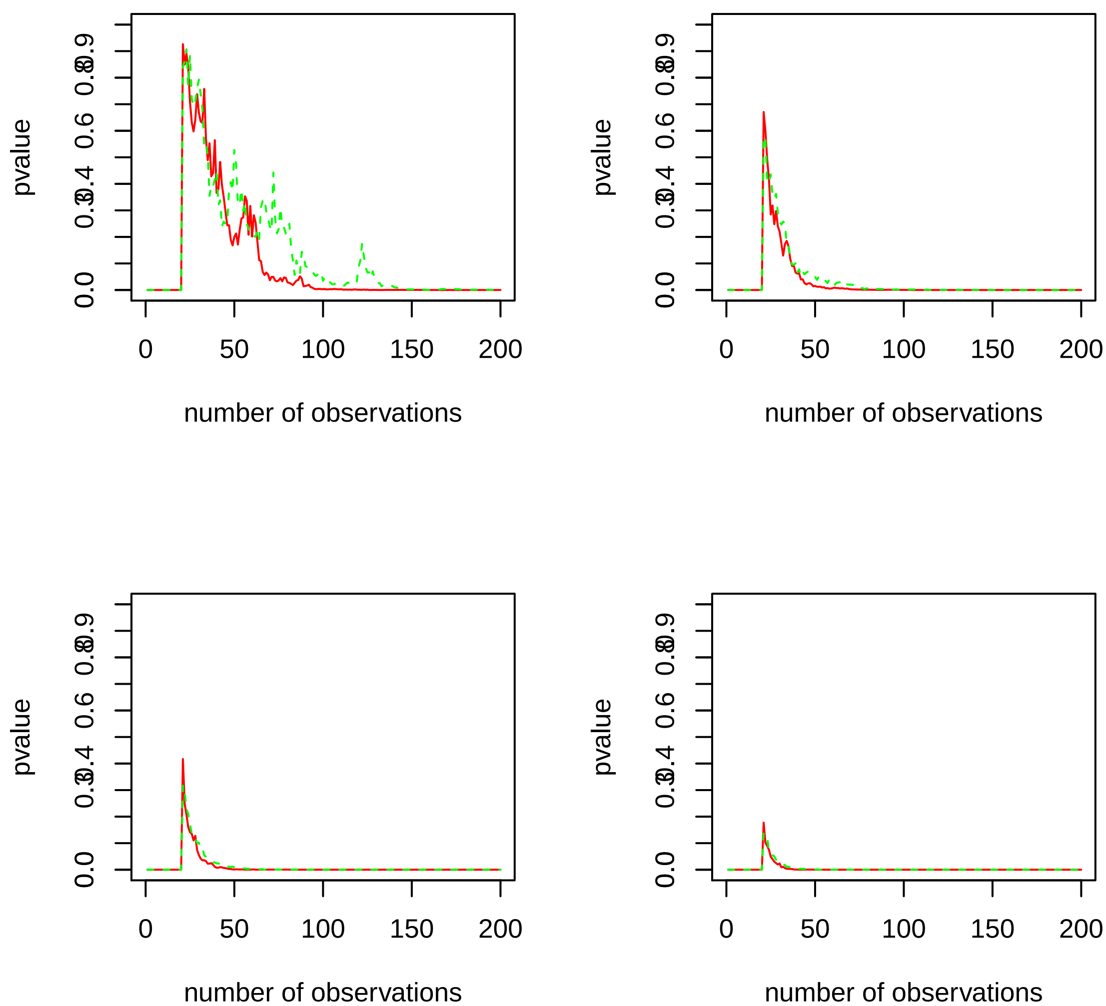


Figure 3.11: P-value Comparison. $\tau = 20$. Change in Gumbel distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1)$ to $(-0.5, 1)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Gumbel likelihood ratio procedure and normal likelihood ratio procedure. The Gumbel likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

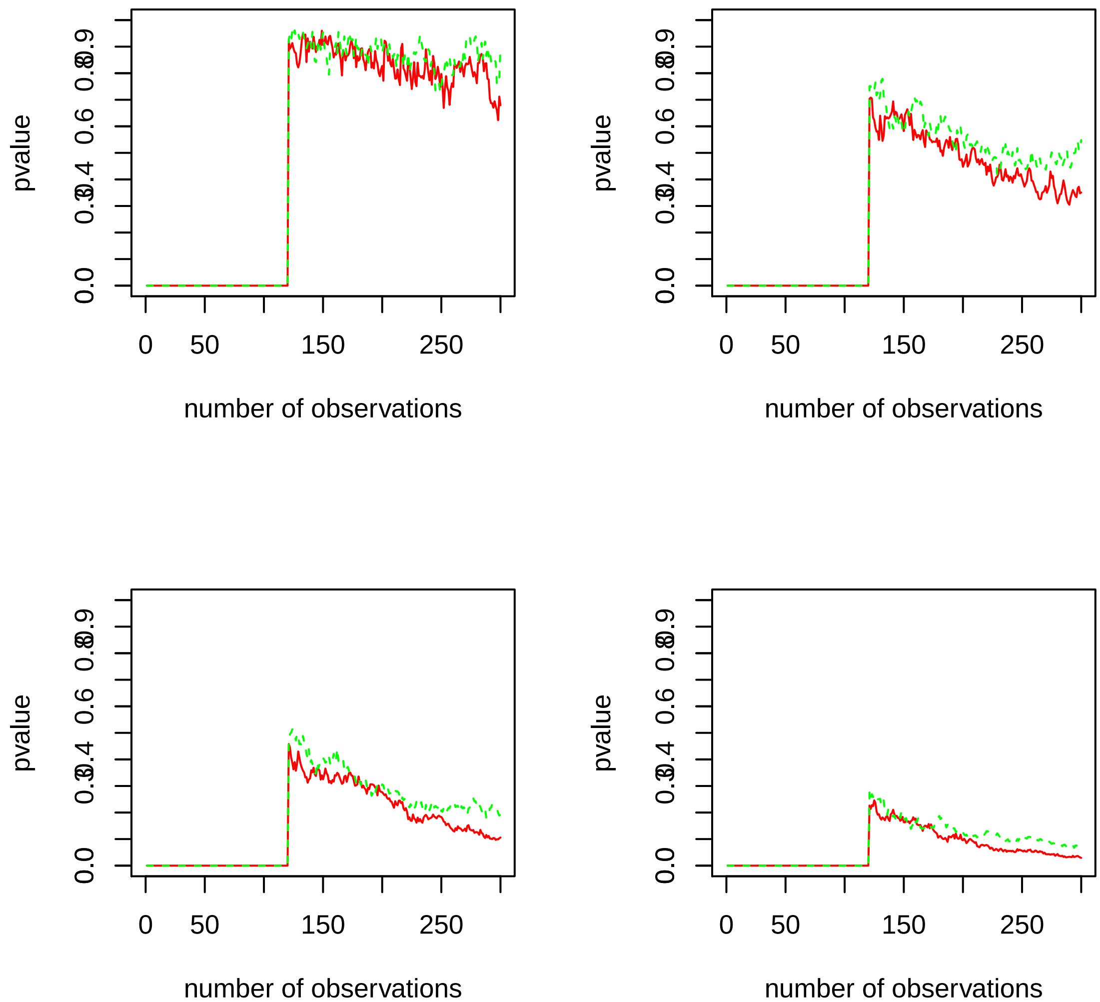


Figure 3.12: P-value Comparison. $\tau = 120$. Change in Gumbel distribution with parameter change from $(\mu, \sigma) = (0, 1)$ to $(0.1, 1)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Gumbel likelihood ratio procedure and normal likelihood ratio procedure. The Gumbel likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

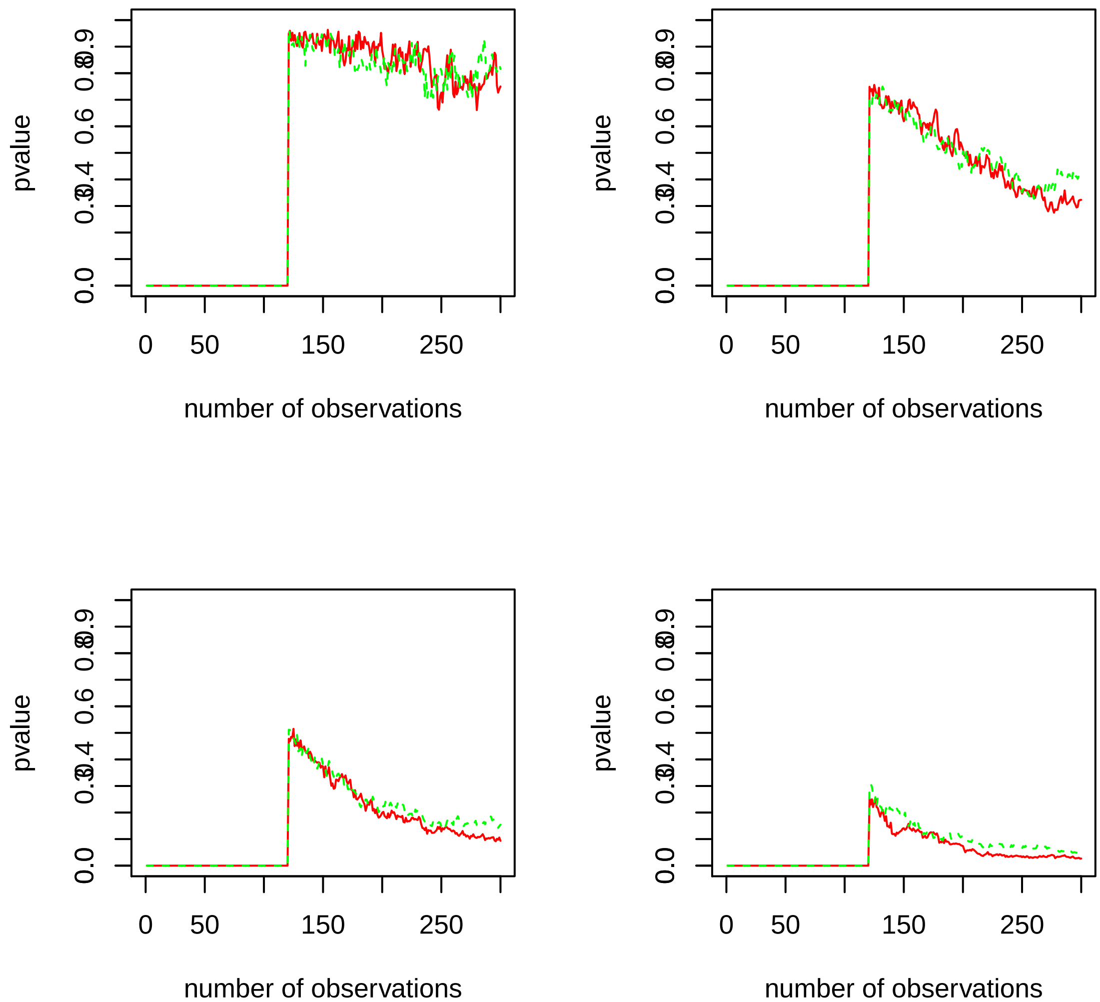


Figure 3.13: P-value Comparison. $\tau = 120$. Change in Gumbel distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1)$ to $(-0.1, 1)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Gumbel likelihood ratio procedure and normal likelihood ratio procedure. The Gumbel likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

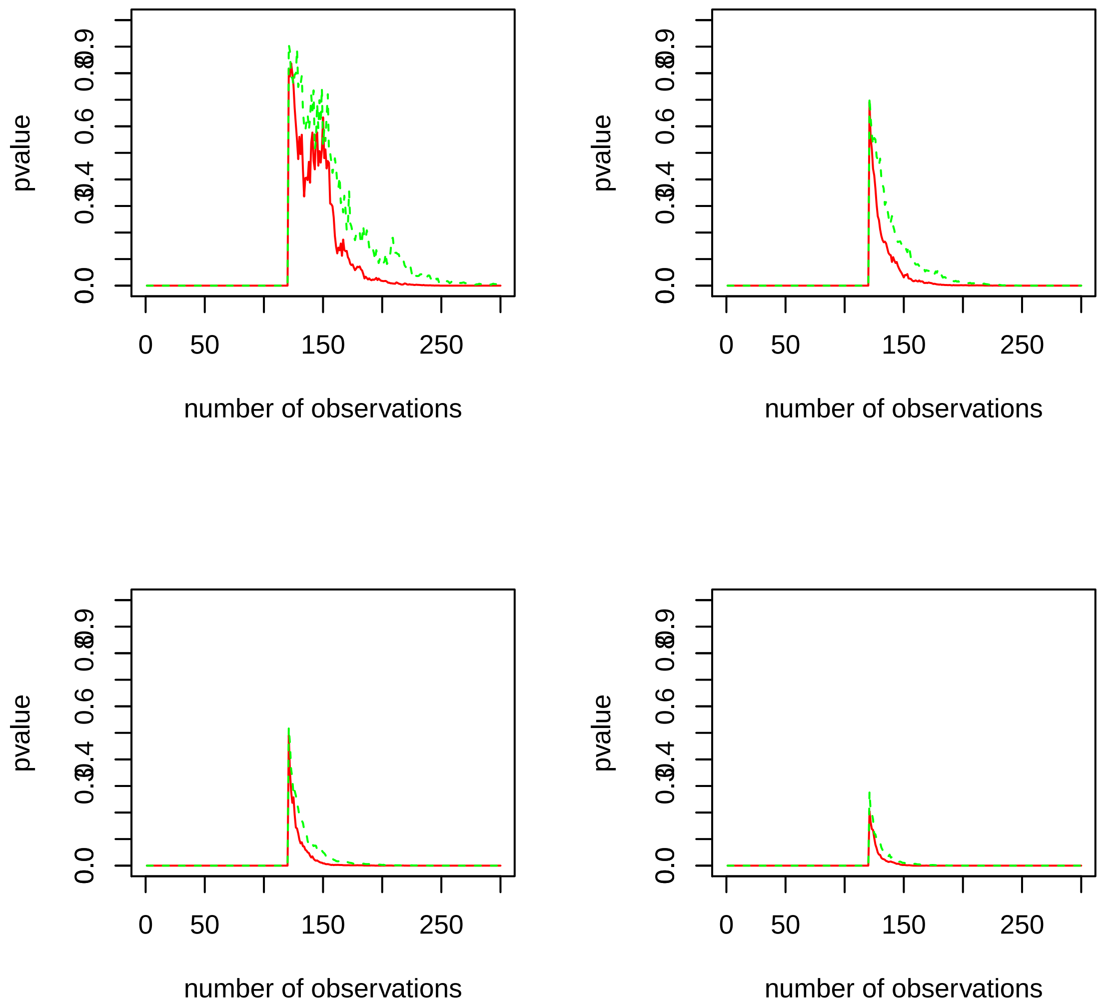


Figure 3.14: P-value Comparison. $\tau = 120$. Change in Gumbel distribution with parameter change from $(\mu, \sigma) = (0, 1)$ to $(0.5, 1)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Gumbel likelihood ratio procedure and normal likelihood ratio procedure. The Gumbel likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

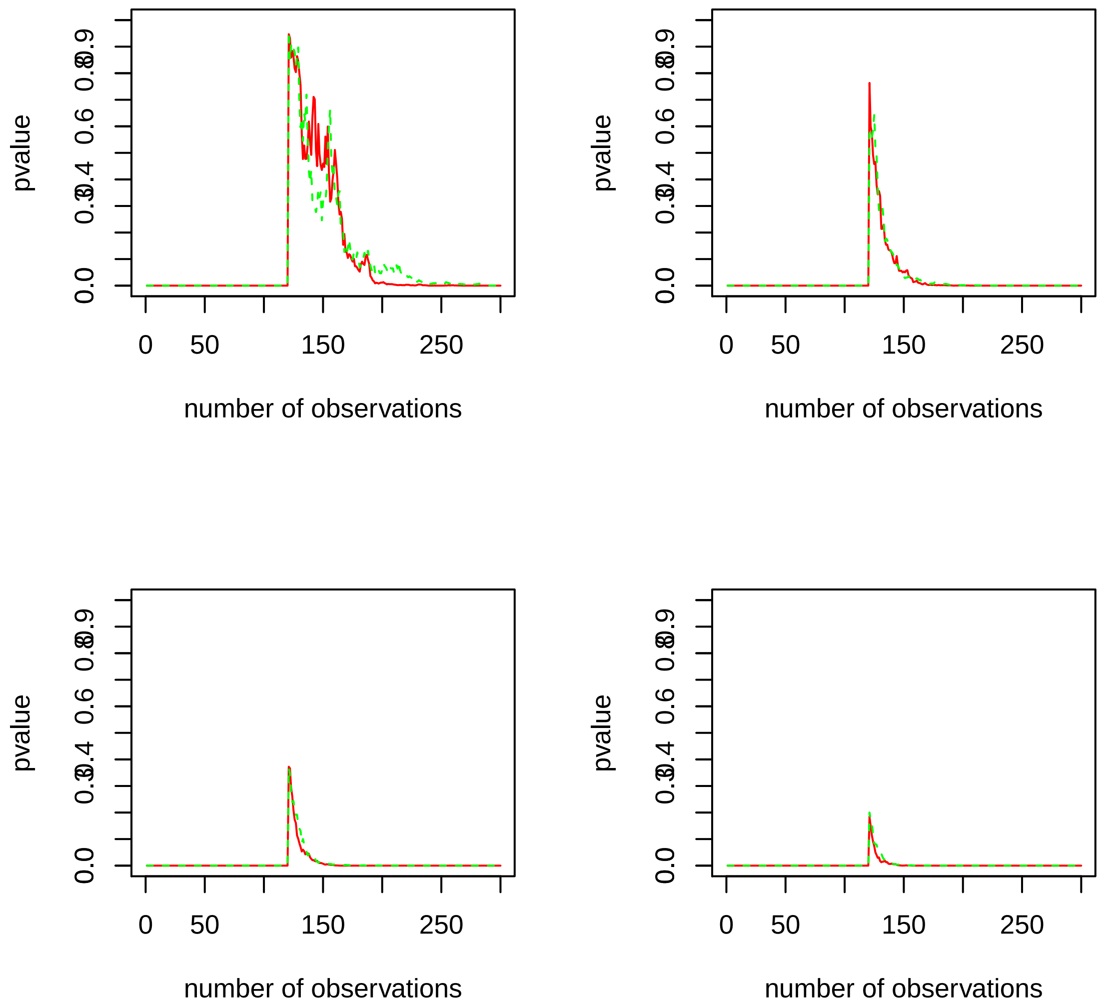


Figure 3.15: P-value Comparison. $\tau = 120$. Change in Gumbel distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1)$ to $(-0.5, 1)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Gumbel likelihood ratio procedure and normal likelihood ratio procedure. The Gumbel likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

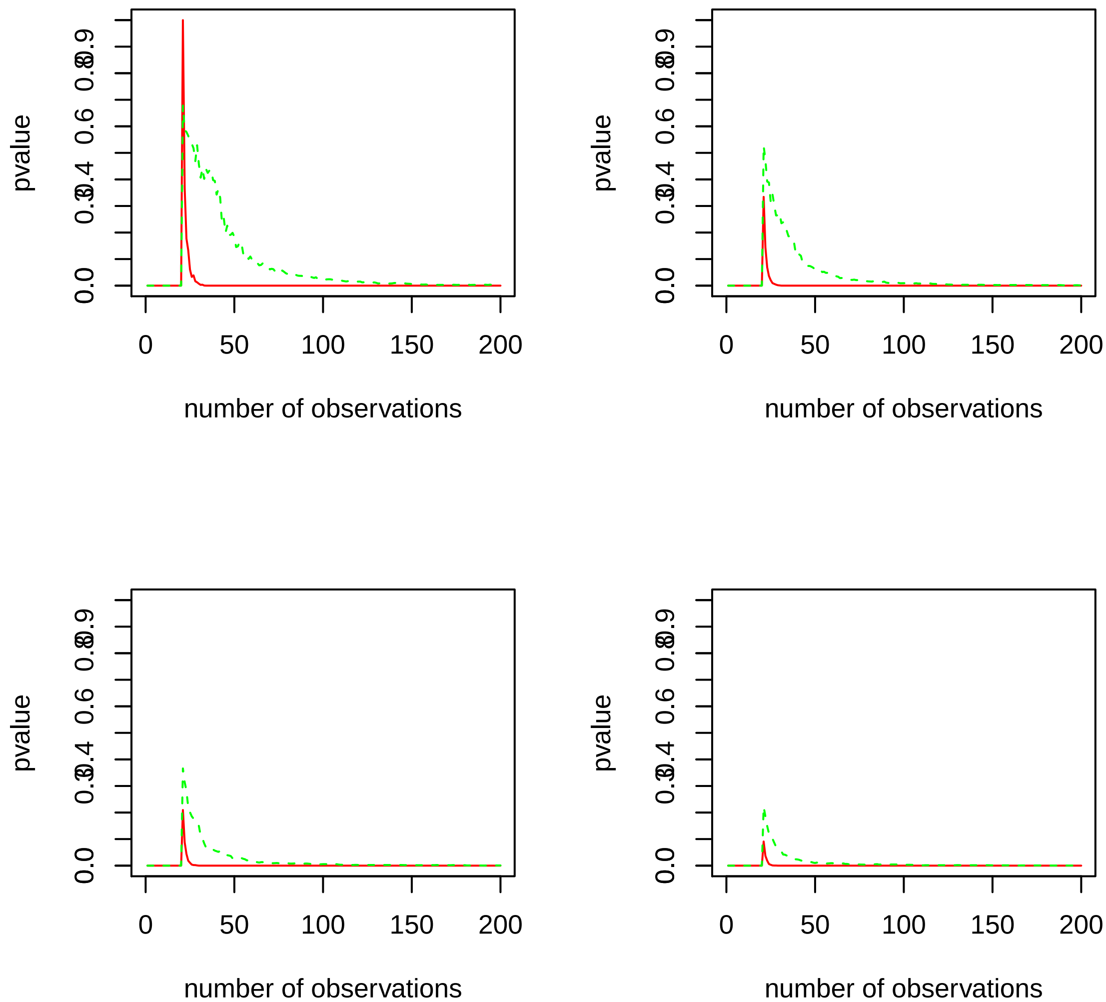


Figure 3.16: P-value Comparison. $\tau = 20$. Change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(0.5, 1, 3)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Fréchet likelihood ratio procedure versus the normal likelihood ratio procedure. The Fréchet likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

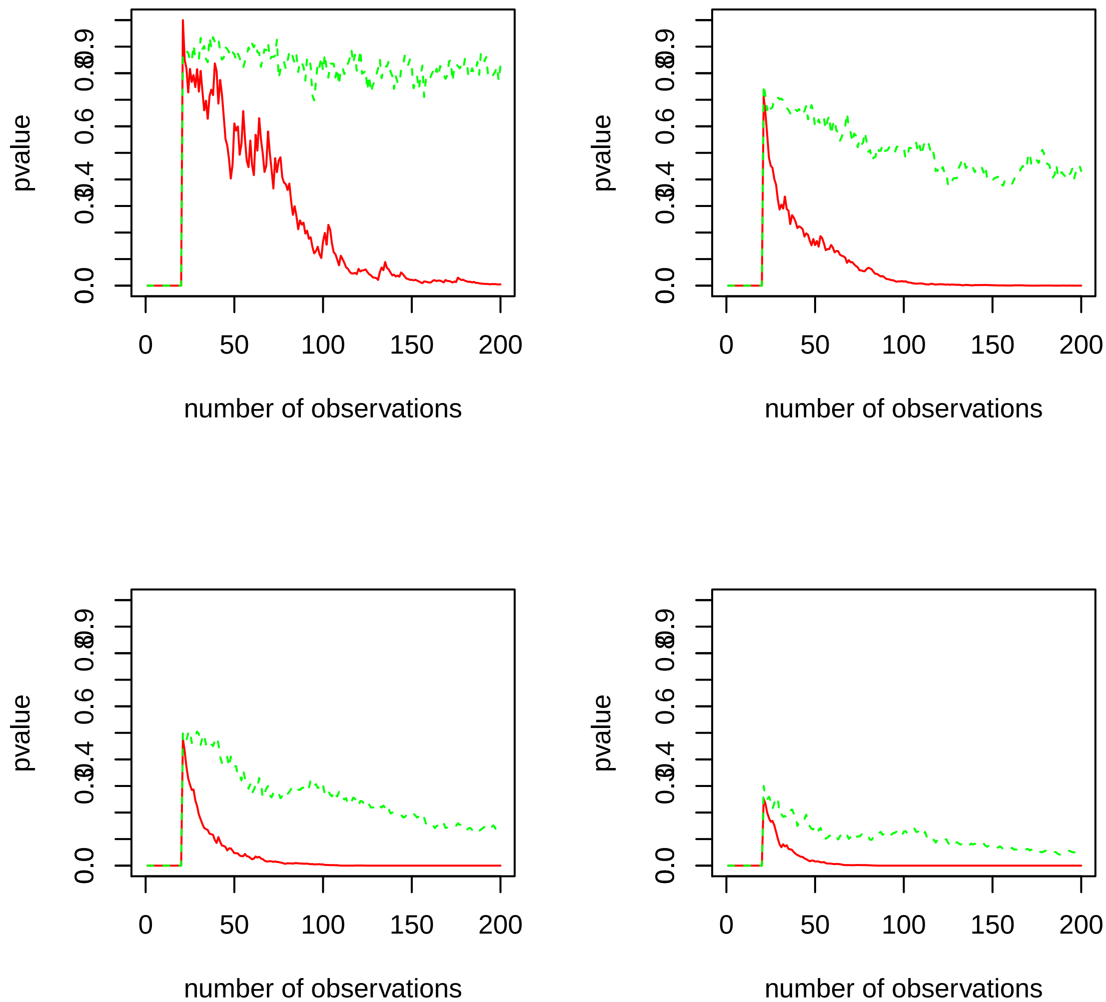


Figure 3.17: P-value Comparison. $\tau = 20$. Change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(0.1, 1, 3)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Fréchet likelihood ratio procedure versus the normal likelihood ratio procedure. The Fréchet likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

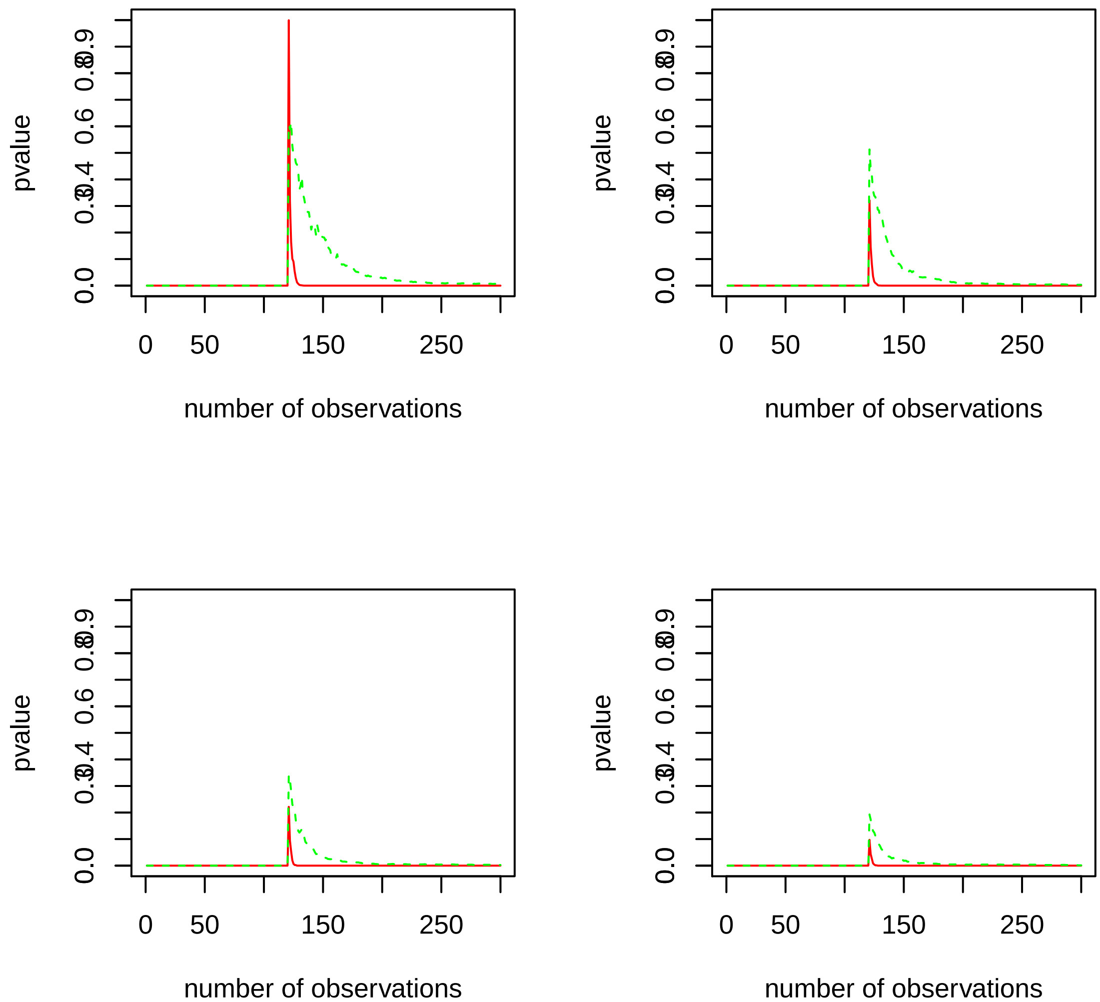


Figure 3.18: P-value Comparison. $\tau = 120$. Change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(0.5, 1, 3)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Fréchet likelihood ratio procedure versus the normal likelihood ratio procedure. The Fréchet likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

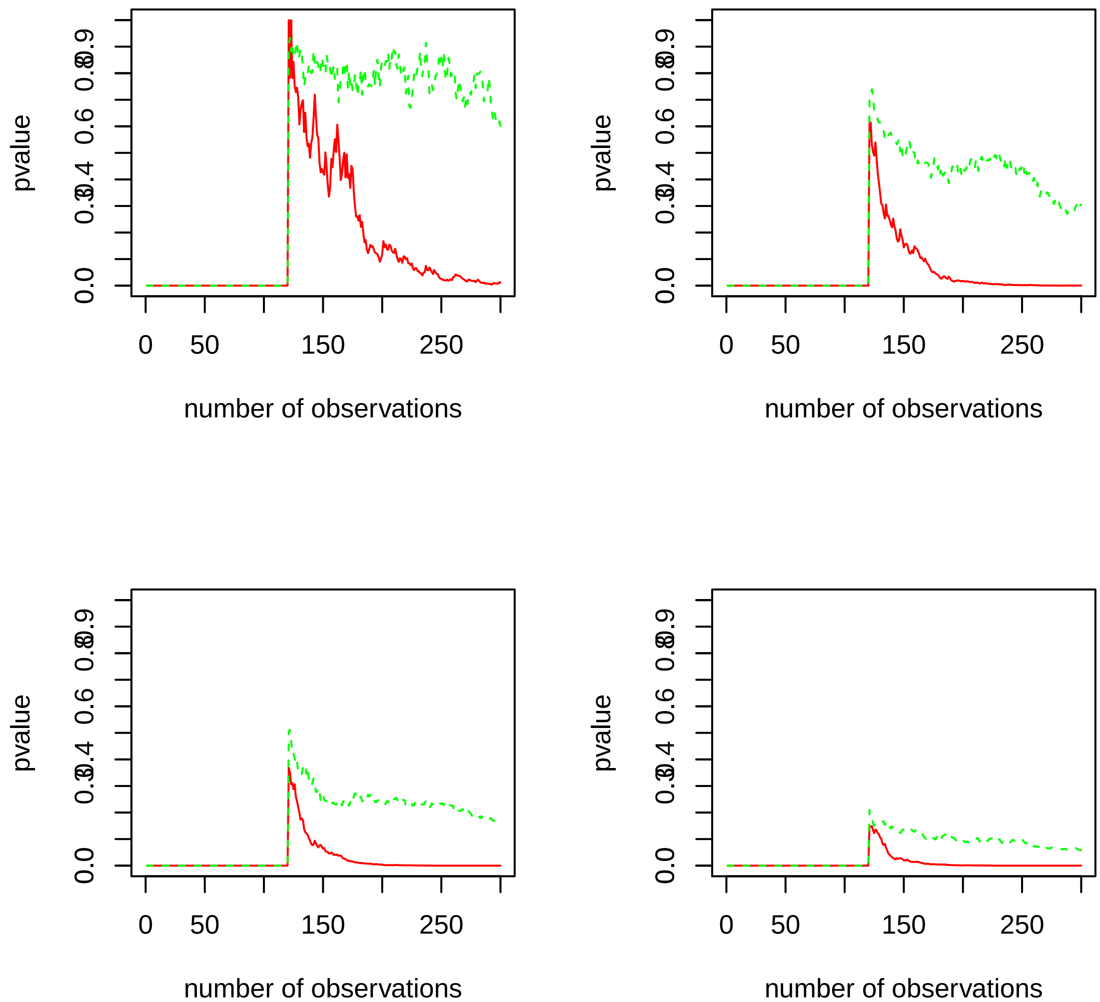


Figure 3.19: P-value Comparison. $\tau = 120$. Change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(0.1, 1, 3)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Fréchet likelihood ratio procedure versus the normal likelihood ratio procedure. The Fréchet likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

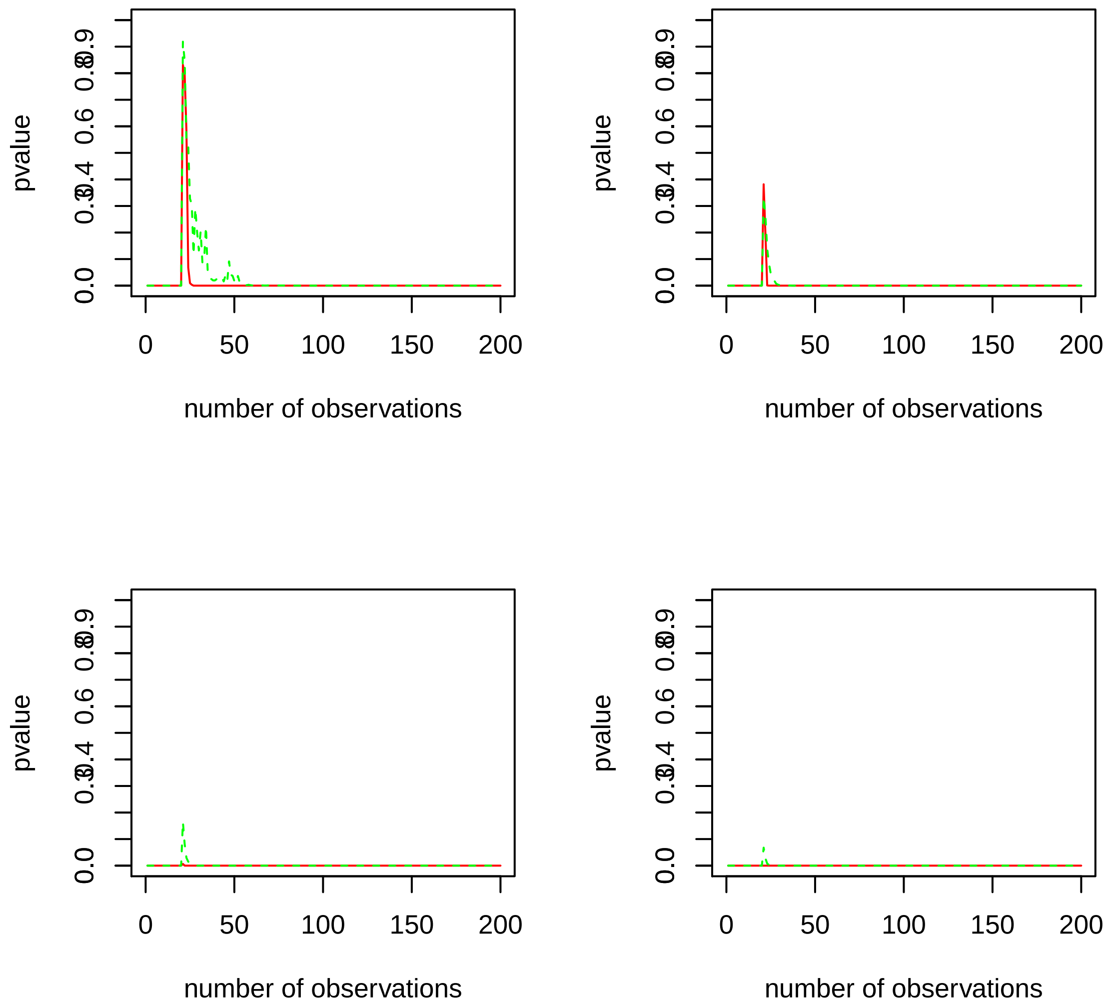


Figure 3.20: P-value Comparison. $\tau = 20$. Change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(-0.5, 1, 3)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Fréchet likelihood ratio procedure versus the normal likelihood ratio procedure. The Fréchet likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

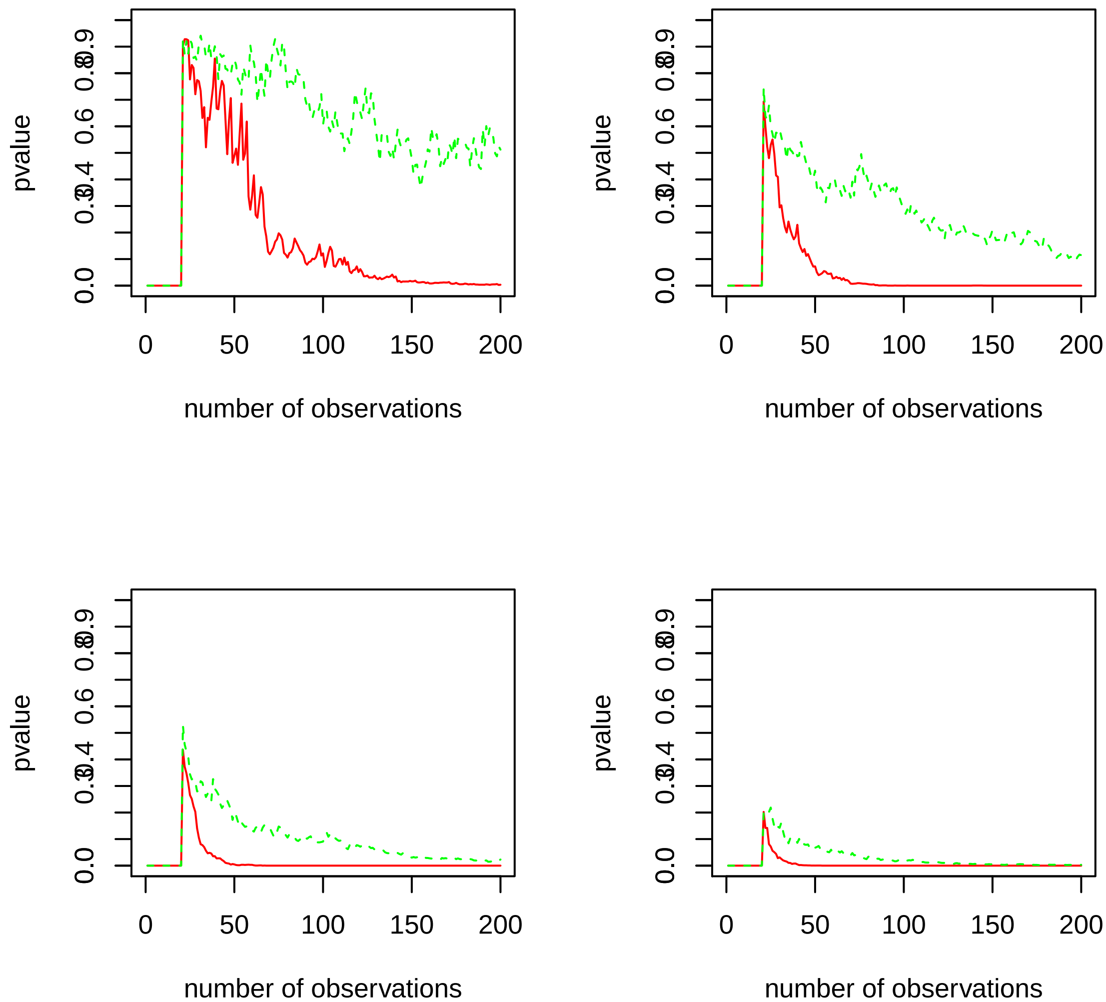


Figure 3.21: P-value Comparison. $\tau = 20$. Change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(-0.1, 1, 3)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Fréchet likelihood ratio procedure versus the normal likelihood ratio procedure. The Fréchet likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

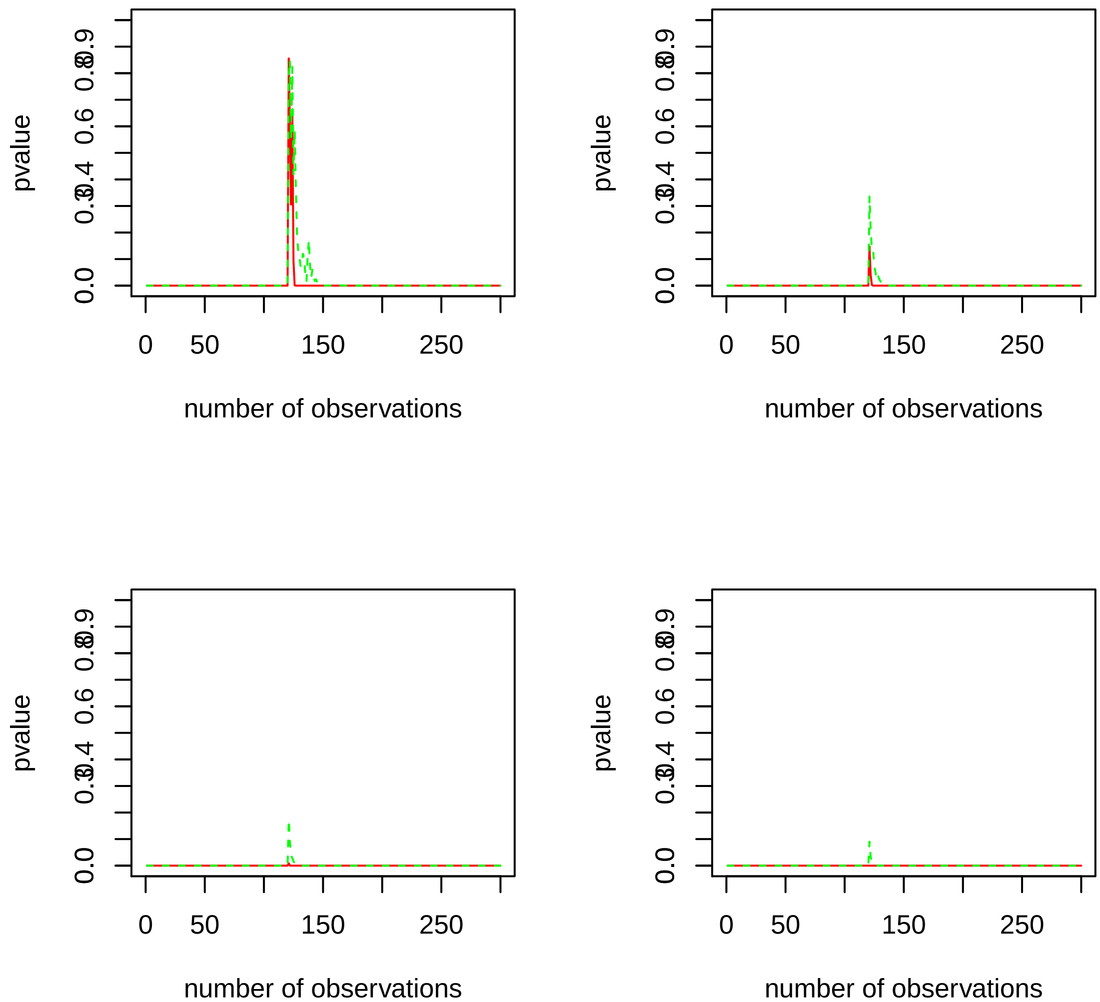


Figure 3.22: P-value Comparison. $\tau = 120$. Change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(-0.5, 1, 3)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Fréchet likelihood ratio procedure versus the normal likelihood ratio procedure. The Fréchet likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

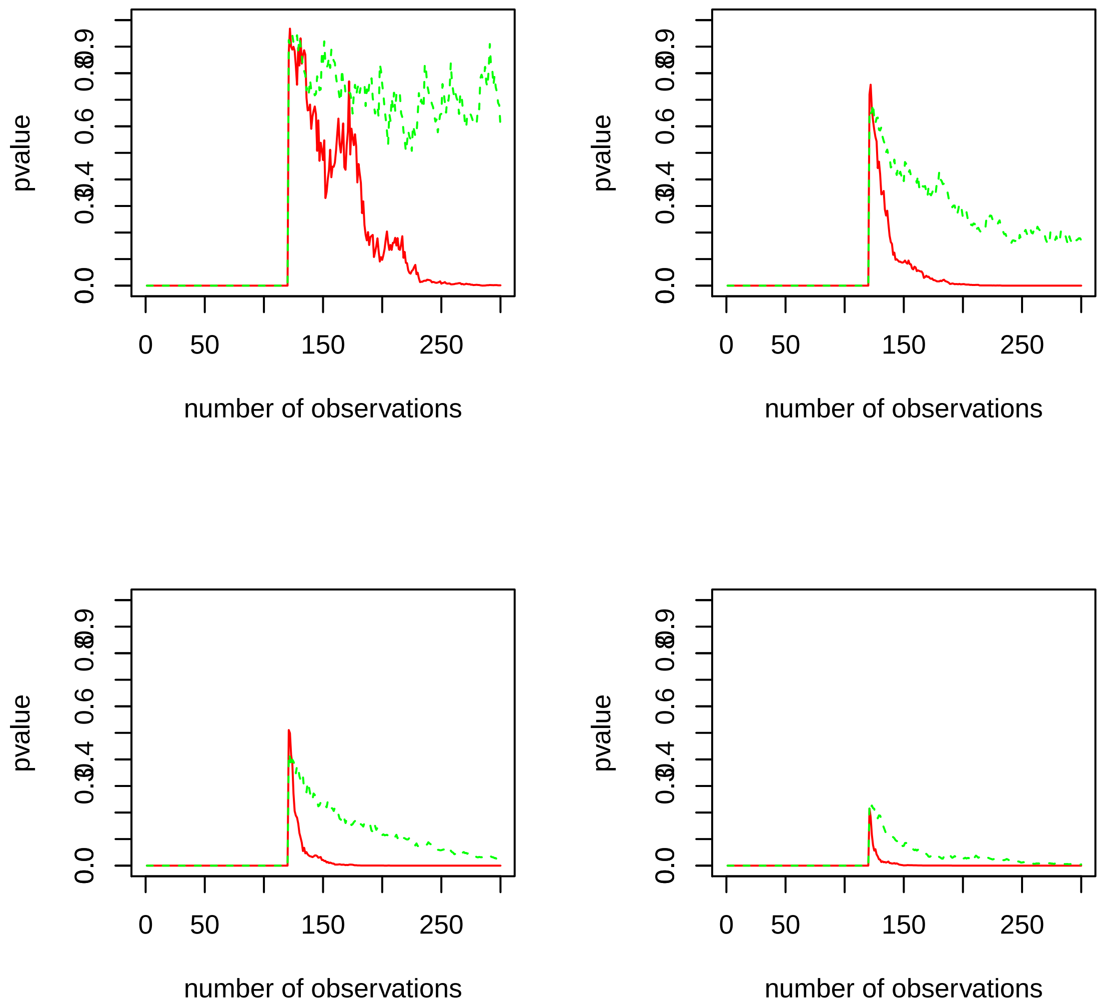


Figure 3.23: P-value Comparison. $\tau = 120$. Change in Fréchet distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1, 3)$ to $(-0.1, 1, 3)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Fréchet likelihood ratio procedure versus the normal likelihood ratio procedure. The Fréchet likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

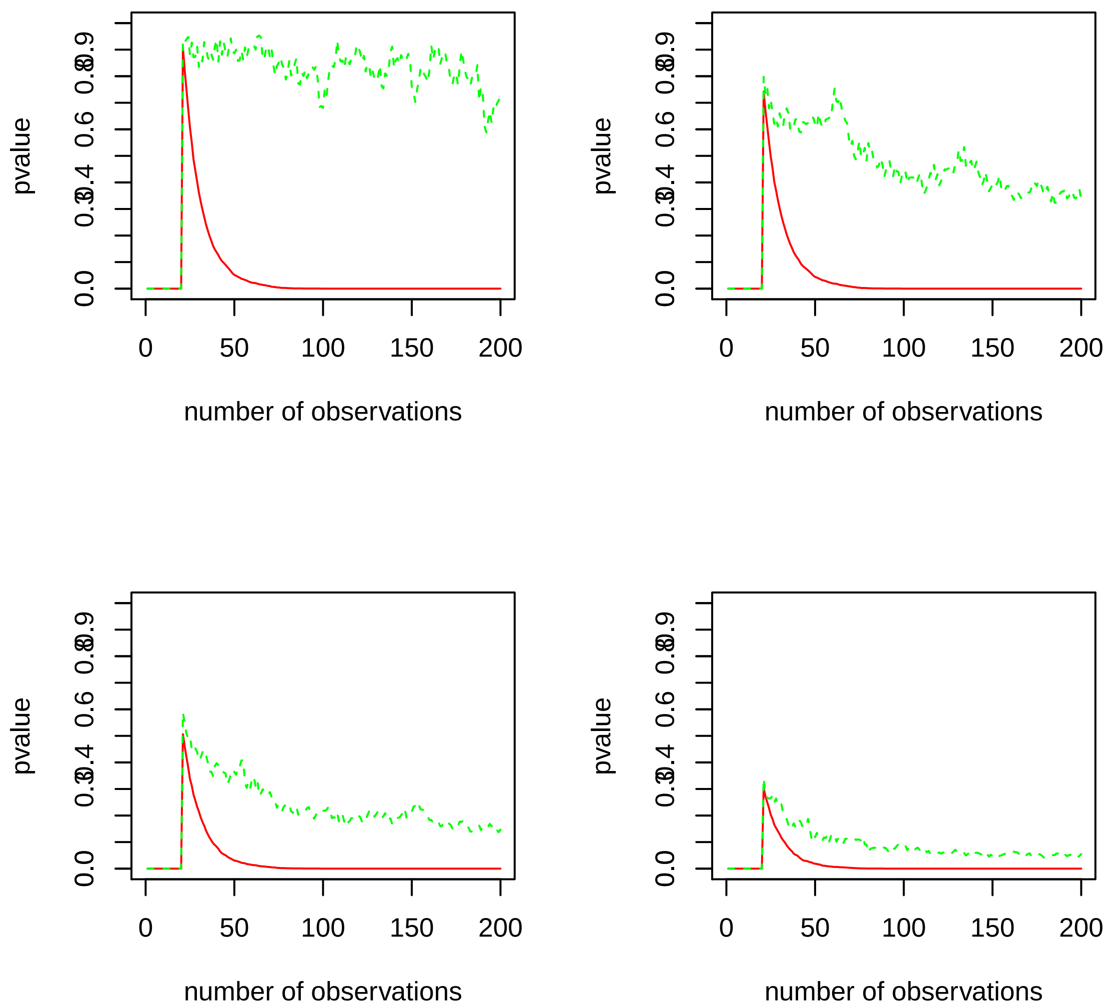


Figure 3.24: P-value Comparison. $\tau = 20$. Change in Generalized Pareto distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1) \to (0.1, 1)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Generalized Pareto likelihood ratio procedure versus the normal likelihood ratio procedure. The Generalized Pareto likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

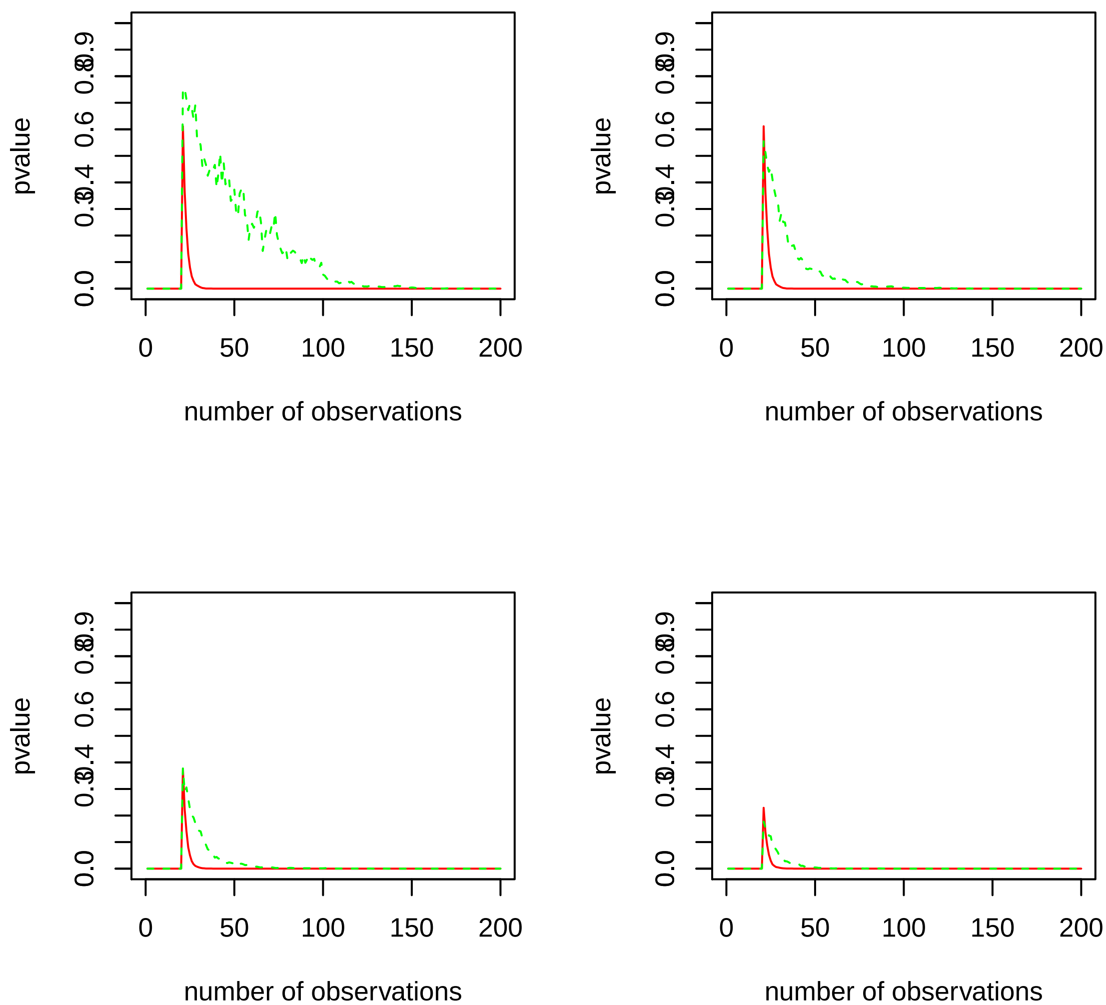


Figure 3.25: P-value Comparison. $\tau = 20$. Change in Generalized Pareto distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1) \rightarrow (0.5, 1)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Generalized Pareto likelihood ratio procedure versus the normal likelihood ratio procedure. The Generalized Pareto likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

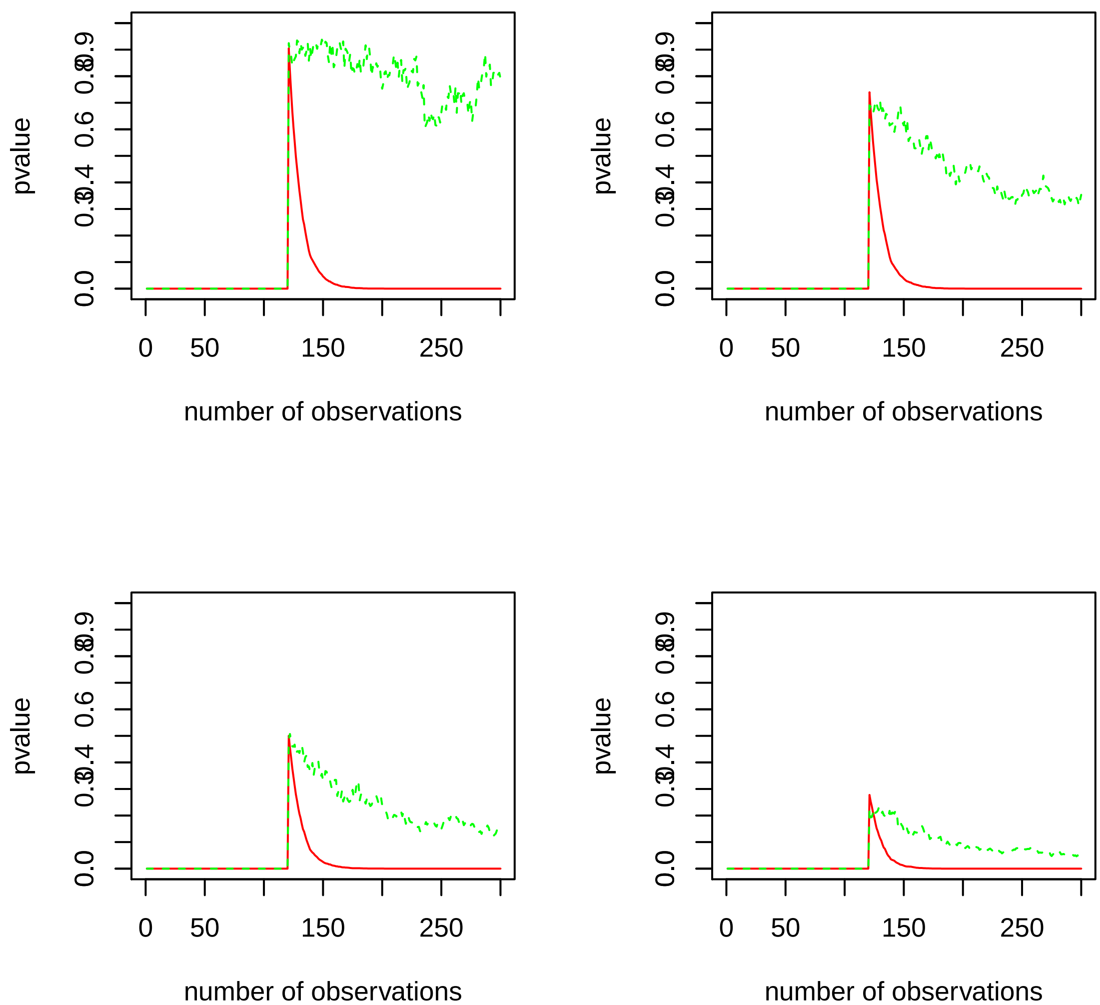


Figure 3.26: P-value Comparison. $\tau = 120$. Change in Generalized Pareto distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1) \rightarrow (0.1, 1)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Generalized Pareto likelihood ratio procedure versus the normal likelihood ratio procedure. The Generalized Pareto likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

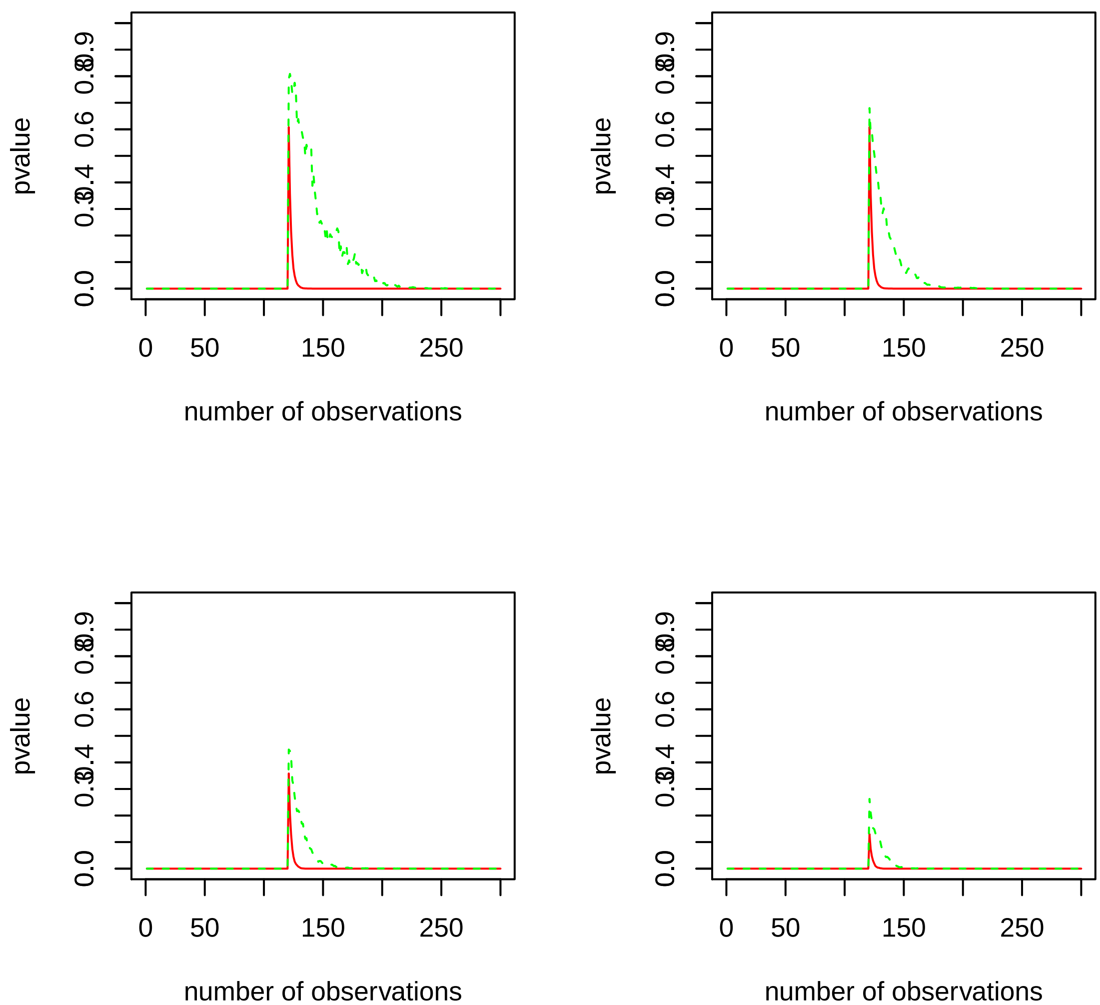


Figure 3.27: P-value Comparison. $\tau = 120$. Change in Generalized Pareto distribution with parameter change from $(\mu, \sigma, \alpha) = (0, 1) \rightarrow (0.5, 1)$. The four panels show 95% (topleft), 75% (topright), 50% (bottomleft)and 25% (bottomright) of p values for Generalized Pareto likelihood ratio procedure versus the normal likelihood ratio procedure. The Generalized Pareto likelihood ratio procedure is marked red and the Normal likelihood ratio procedure is marked green.

Distribution	$c_1 = \frac{1}{4}$	$c_1 = \frac{1}{2}$	$c_1 = 1$
$c_2 = 1$	1927	1922	1918
$c_2 = 1.5$	1906	1906	1906
$c_2 = 2$	1906	1906	1906

Table 3.1: Average Run Length Approach: Atlantic hurricane data from 1851 to 2008 are used to detect the distribution change of the maximum sustained winds, assuming no change between 1851-1880.

Distribution	$c_1 = \frac{1}{4}$	$c_1 = \frac{1}{2}$	$c_1 = 1$
$c_2 = 1$	1885	1885	1886
$c_2 = 1.5$	1906	1906	1899
$c_2 = 2$	1894	1906	1906

Table 3.2: P-value Approach: Atlantic hurricane data from 1851 to 2008 are used to detect the distribution change in the maximum sustained winds, assuming no change between 1851-1880.

Chapter 4

Detecting Change in Extreme Order Statistic

This chapter extends the framework in Chapter 3 and discusses the change detection problem of extreme data with order statistic. Section 4.1 presents an overview of the problem and the procedure we will follow. In Section 4.2, a general detection methodology is presented in the context of employing order statistic before details are revealed in Section 4.3 and 4.4. Section 4.3 discusses generating order statistic using the rejection sampling method, which is an extension of generating the maxima (minima). Section 4.4 discusses how to compute the maximum likelihood estimation for the generalized extreme value distribution parameters. The main idea is to transform this problem into a linearly constrained optimization problem. Section 4.5 uses the hurricane example to illustrate our methodology.

4.1 Problem Formulation

Assume we have m blocks of data. In each block i, we observe $z_i^{(1)}, z_i^{(2)}, \ldots, z_i^{(r)}$, where $z_i^{(1)} \geq z_i^{(2)}, \ldots, \geq z_i^{(r)}$ for $i = 1, 2, \ldots, m$. The joint probability density function for $z_i^{(1)}, z_i^{(2)}, \ldots, z_i^{(r)}$ is

$$L(z_i^{(1)}, z_i^{(2)}, \ldots, z_i^{(r)}) = \exp\{-\{1 + \xi \frac{z_i^{(r)} - \mu}{\sigma}\}^{-1/\xi}\} \prod_{k=1}^{r} \sigma^{-1}\{1 + \xi \frac{z_i^{(k)} - \mu}{\sigma}\}^{-1-1/\xi} \quad (4.1.1)$$

given $1 + \xi \frac{z_i^{(k)} - \mu}{\sigma} > 0$ for $k = 1, 2, \ldots, r$.

Therefore, the likelihood function is

$$L(z_i^{(1)}, z_i^{(2)}, \ldots, z_i^{(r)}, i = 1, 2, \ldots, m) \tag{4.1.2}$$

$$= \prod_{i=1}^{\tau} \exp\{-\{1 + \xi_1 \frac{z_i^{(r)} - \mu_1}{\sigma_1}\}^{-1/\xi_1}\} \prod_{k=1}^{r} \sigma_1^{-1}\{1 + \xi_1 \frac{z_i^{(k)} - \mu_1}{\sigma_1}\}^{-1-1/\xi_1} \tag{4.1.3}$$

$$\times \prod_{i=\tau+1}^{m} \exp\{-\{1 + \xi_2 \frac{z_i^{(r)} - \mu_2}{\sigma_2}\}^{-1/\xi_2}\} \prod_{k=1}^{r} \sigma_2^{-1}\{1 + \xi_2 \frac{z_i^{(k)} - \mu_2}{\sigma_2}\}^{-1-1/\xi_2} \tag{4.1.4}$$

given $1 + \xi \frac{z_i^{(k)} - \mu}{\sigma} > 0$ for $k = 1, 2, \ldots, r$ and $i = 1, 2, ...m$.

The change detection problem can be translated to a hypothesis testing problem: $H_0 : \tau \geq m$ $H_1 : \tau < m$. H_0 indicates no change in the whole series, whereas H_1 indicates a change.

Since we have no idea about (μ_1, σ_1, ξ_1), (μ_2, σ_2, ξ_2) and τ, it is safe to apply the maximum likelihood approach. In theory, τ can be any value as long as it satisfies $0 \leq \tau \leq m$, but because we need to estimate (μ_1, σ_1, ξ_1) and (μ_2, σ_2, ξ_2), if τ is too close to the beginning or the end of the data series, the maximum likelihood estimation for either (μ_1, σ_1, ξ_1) or (μ_2, σ_2, ξ_2) will be poor and has high variance, which is detrimental to the analysis in the later stage. For this purpose, we limit $t_0 \leq \tau \leq t_1$, where t_0 and t_1 are both positive integers. This ensures sufficient data points for estimating both (μ_1, σ_1, ξ_1) and (μ_2, σ_2, ξ_2). For simplicity, we choose t_0 and t_1 such that t_0 is roughly close to $m - t_1$.

To infer the actual change point τ, which we denote as $\hat{\tau}$, there are two ways to tackle this issue: One way is that for each $t_0 \leq \tau \leq t_1$, obtain $(\hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1)$ from observations $1, 2, \ldots, \tau$, and $(\hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2)$ from observations $\tau + 1, \tau + 2, \ldots, m$ using maximum likelihood. We will explain the methodology to this estimation problem in Section 4.4. We substitute those values into the likelihood function Equation 4.1.2 and compute their likelihood values. Finally, we choose $\hat{\tau}$ and the corresponding $(\hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1)$ and $(\hat{\mu}_1, \hat{\sigma}_2, \hat{\xi}_2)$ such that the likelihood function of Equation 4.1.2 is maximized.

Another approach is to find out which $\hat{\tau}$ best separates the parameter vectors (μ_1, σ_1, ξ_1) and (μ_2, σ_2, ξ_2), which is the methodology used in Chapter 4. For each τ, we obtain $(\hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1)$ from observations $1, 2, \ldots, \tau$ and $(\hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2)$ from observations $\tau + 1, \ldots, m$ using the maximum likelihood approach as usual. We wish to find a $\hat{\tau}$ such that the distance between $(\hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1)$ and $(\hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2)$ is maximized. Here we use the L_2 norm as a distance measure, which is expressed as $||x|| = \sqrt{x_1^2 + \ldots + x_p^2}$, where p is the dimension of the vector $\mathbf{x}$.

After obtaining $\hat{\tau}$ and the corresponding $(\hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1)$ and $(\hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2)$, we record

$$\delta = (\hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1) - (\hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2)$$

This serves as our "naive" test statistic.

Our next job is to simulate a null setting, the key of which being the choice of the null parameter. Based on the location of $\hat{\tau}$, we obtain the estimated null parameter θ_0^*, with which we use to generate random sample from the generalized extreme value distribution of size m. To determine θ_0^*, if $\hat{\tau} > \frac{t_0+t_1}{2}$, we obtain θ_0^* based on the observations $1, \ldots, \hat{\tau} - k_0$. Otherwise, we obtain θ_0^* based on the observations $\hat{\tau} + k_0, \ldots, m$, where k_0 is our buffer, and typically a small number.

The intention of estimating θ_0^* and introducing another parameter k_0 reflects our conservative approach: if $\hat{\tau} > \frac{t_0+t_1}{2}$, there are sufficient observations in the first half, which generally leads to a better estimate. Since $\hat{\tau}$ is our estimation and in most cases, might not be exactly the true change point τ, we bring in errors. To allow the parameter k_0 into the analysis, we want to mitigate side effect of the estimation error of τ. Leveraging observations $1, \ldots, \hat{\tau} - k_0$ when $\hat{\tau}$ is large and observations $\hat{\tau} + k_0, \ldots, m$ when $\hat{\tau}$ is small contributes more accuracy in the estimation of θ_0^*.

In simulating the null setting, we need to generate independent and identically distributed random samples from θ_0^* in order to perform parametric bootstrap. Section 4.3 explains in detail how to use rejection sampling to simulate order statistic in its vector form. For each simulation $b = 1, 2, \ldots, B$, repeat the previous steps in finding $\hat{\tau}_b$, and $(\hat{\mu}_{1b}, \hat{\sigma}_{1b}, \hat{\xi}_{1b})$ and $(\hat{\mu}_{2b}, \hat{\sigma}_{2b}, \hat{\xi}_{2b})$, and compute $\delta_b = (\hat{\mu}_{1b}, \hat{\sigma}_{1b}, \hat{\xi}_{1b}) - (\hat{\mu}_{2b}, \hat{\sigma}_{2b}, \hat{\xi}_{2b})$.

Our final step is to compute the p-value of δ_b against the null distribution formed by δ_b, $b = 1, 2, \ldots, B$. This can be done in several ways including multivariate quantile techniques and standardization. Here we adopt the latter approach by standardizing our "naive" test statistic δ_b. From δ_b, $b = 1, 2, \ldots, B$, it is easy to obtain the sample variance-covariance matrix $\hat{\Sigma}$, which is defined as:

$$\hat{\Sigma} = \frac{1}{B-1} \sum_{b=1}^{B} (\delta_b - \bar{\delta})(\delta_b - \bar{\delta})'$$

where $\bar{\delta} = \frac{1}{B} \sum_{b=1}^{B} \delta_b$.

Our test statistic is defined as $R = \delta \hat{\Sigma}^{-1} \delta'$, and the null test statistics are defined as $R_b = \delta_b \hat{\Sigma}^{-1} \delta_b'$ for $b = 1, 2, \ldots, B$. To avoid the situation that the sample variance-covariance matrix might have a large conditional number, we make a small adjustment by defining $\tilde{R} = \delta(\hat{\Sigma} + \Lambda)^{-1}\delta'$, and define the null test statistics as $\tilde{R}_b = \delta_b(\hat{\Sigma} + \Lambda)^{-1}\delta_b'$, where $\Lambda = diag\{\epsilon, \ldots, \epsilon\}$. Here ϵ is a very small number, say 10^{-6} in our analysis.

To compute the p-value, we simply compute the quantile of $\tilde{R}$ against $\tilde{R}_b$, where $b = 1, 2, \ldots, B$. If no change model fits the real data, we should expect to see $\tilde{R}$ in the middle range in the empirical probability distribution $\hat{F}$ formed by $\tilde{R}_b$, $b = 1, 2, \ldots, B$. On the contrary, if the change model fits the real data, we should expect $\tilde{R}$ to be in the tails of the empirical distribution $\hat{F}$. This p-value gives strength of evidence whether the change model fits the data or not. Small p-value indicates rejection of the no change model and favors the change model.

4.2 Change Detection with Order Statistic

This section provides full details about our procedure.

Assume that the data series are divided into m blocks and for each block i, we consider the largest r_i order statistic together, namely $z_i^{(1)} \geq \ldots \geq z_i^{(r_i)}$. This idea is motivated by [22] on modeling the sea level with order statistic. From our perspective, order statistic can also be used to monitor changes in the extreme behaviors. This is a natural generalization of the situation where $r_i = 1$ for all $i = 1, 2, \ldots, m$. In this section, we aim to generalize the discussion of [15] by establishing a change detection methodology for GEV distribution using order statistic.

Assume we observe $Z_i = (z_i^{(1)}, \ldots, z_i^{(r_i)})$, where $i = 1, 2, \ldots, m$, and $z_i^{(1)}, \ldots, z_i^{(r_i)}$ are the largest r_i order statistics in block i. The likelihood function for block i is

$$L_i(\mu, \xi, \sigma) = \exp\{-\{1 + \xi\frac{z_i^{(r_i)} - \mu}{\sigma}\}^{-1/\xi}\} \times \prod_{k=1}^{r_i} \sigma^{-1}\{1 + \xi\frac{z_i^{(k)} - \mu}{\sigma}\}^{-1/\xi - 1}$$

for $\xi \neq 0$, and $1 + \xi\frac{z_i^{(k)} - \mu}{\sigma} > 0$.

And the likelihood function is

$$L_i(\mu, \sigma, \xi) = \exp\{-\exp\{-\frac{z_i^{(r_i)} - \mu}{\sigma}\}\} \times \prod_{k=1}^{r_i} \sigma^{-1}\exp\{-\frac{z_i^{(k)} - \mu}{\sigma}\}$$

for $\xi = 0$.

We are interested in whether there is any change in block τ and onwards, where $0 \leq \tau \leq m$. To formulate it into a hypothesis testing problem, we test $H_0 : \tau \geq m$ VS $H_1 : 0 \leq \tau < m$. First let us consider the change from μ to $\mu + \delta$ under $\xi \neq 0$.
The likelihood function for the null hypothesis is

$$L_0(\mu, \xi, \sigma) = \prod_{i=1}^{m} \{\exp\{-\{1 + \xi \frac{z_i^{(r_i)} - \mu}{\sigma}\}^{-1/\xi}\} \times \prod_{k=1}^{r_i} \sigma^{-1}\{1 + \xi \frac{z_i^{(k)} - \mu}{\sigma}\}^{-1/\xi - 1}\}$$

provided that $1 + \xi \frac{z_i^{(k)} - \mu}{\sigma} > 0$ for $k = 1, 2..., r_i$ and $i = 1, 2..., m$.

Under the alternative, the likelihood function is

$$\begin{aligned} L_1(\mu, \xi, \sigma) &= \prod_{i=1}^{\tau} \{\exp\{-\{1 + \xi \frac{z_i^{(r_i)} - \mu}{\sigma}\}^{-1/\xi}\} \times \prod_{k=1}^{r_i} \sigma^{-1}\{1 + \xi \frac{z_i^{(k)} - \mu}{\sigma}\}^{-1/\xi - 1}\} \\ &\times \prod_{i=\tau+1}^{m} \{\exp\{-\{1 + \xi \frac{z_i^{(r_i)} - \mu - \delta}{\sigma}\}^{-1/\xi}\} \times \prod_{k=1}^{r_i} \sigma^{-1}\{1 + \xi \frac{z_i^{(k)} - \mu - \delta}{\sigma}\}^{-1/\xi - 1}\} \end{aligned}$$

provided that $1 + \xi \frac{z_i^{(k)} - \mu}{\sigma} > 0$ for $i = 1, 2, \ldots, \tau$ and $1 + \xi \frac{z_i^{(k)} - \mu - \delta}{\sigma} > 0$ for $i = \tau + 1, \ldots, m$. $k = 1, 2, \ldots, r_i$.

If $\exists j$, such that $\min(1+\xi \frac{z_j^{(1)} - \mu}{\sigma}, 1+\xi \frac{z_j^{(r_j)} - \mu}{\sigma}) \leq 0$, we reject the null because the likelihood for the null will be zero. Otherwise, we assume that $\forall j$, $\min(1+\xi \frac{z_j^{(1)} - \mu}{\sigma}, 1+\xi \frac{z_j^{(r_j)} - \mu}{\sigma}) > 0$ and $k_0 = \inf\{t : \min(1 + \xi \frac{z_{t+1}^{(1)} - \mu - \delta}{\sigma}, 1 + \xi \frac{z_{t+1}^{(r_{t+1})} - \mu - \delta}{\sigma}), \min(1 + \xi \frac{z_m^{(1)} - \mu - \delta}{\sigma}, 1 + \xi \frac{z_m^{(r_m)} - \mu - \delta}{\sigma}) > 0\}$.

Clearly, $\tau \geq k_0$.

$$
\begin{aligned}
& \Lambda(\mu,\xi,\sigma) = \frac{\max_{k_0\leq\tau<m} L_1(\tau)}{\max_{\tau\geq m} L_0(\tau)} \\
= \;& \{\prod_{i=1}^{m}\{\exp\{-\{1+\xi\frac{z_i^{(r_i)}-\mu}{\sigma}\}^{-1/\xi}\} \times \prod_{k=1}^{r_i}\sigma^{-1}\{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}\}^{-1/\xi-1}\}\}^{-1} \\
\times \;& \prod_{i=1}^{\tau}\{\exp\{-\{1+\xi\frac{z_i^{(r_i)}-\mu}{\sigma}\}^{-1/\xi}\} \times \prod_{k=1}^{r_i}\sigma^{-1}\{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}\}^{-1/\xi-1}\} \\
\times \;& \prod_{i=\tau+1}^{m}\{\exp\{-\{1+\xi\frac{z_i^{(r_i)}-\mu-\delta}{\sigma}\}^{-1/\xi}\} \times \prod_{k=1}^{r_i}\sigma^{-1}\{1+\xi\frac{z_i^{(k)}-\mu-\delta}{\sigma}\}^{-1/\xi-1}\} \\
\propto \;& \prod_{i=1}^{\tau}\{\exp\{-\{1+\xi\frac{z_i^{(r_i)}-\mu}{\sigma}\}^{-1/\xi}\} \times \prod_{k=1}^{r_i}\sigma^{-1}\{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}\}^{-1/\xi-1}\} \\
\times \;& \prod_{i=\tau+1}^{m}\{\exp\{-\{1+\xi\frac{z_i^{(r_i)}-\mu-\delta}{\sigma}\}^{-1/\xi}\} \times \prod_{k=1}^{r_i}\sigma^{-1}\{1+\xi\frac{z_i^{(k)}-\mu-\delta}{\sigma}\}^{-1/\xi-1}\}
\end{aligned}
$$

Maximizing Λ is equivalent to maximizing $\log\Lambda$.

$$
\begin{aligned}
\log\Lambda \;=\;& \sum_{i=1}^{\tau}\{-(1+\xi\frac{z_i^{(r_i)}-\mu}{\sigma})^{-1/\xi} + \sum_{k=1}^{r_i}\{-\log\sigma-(1+1/\xi)\log(1+\xi\frac{z_i^{(k)}-\mu}{\sigma})\}\} \\
+\;& \sum_{i=\tau+1}^{m}\{-(1+\xi\frac{z_i^{(k)}-\mu-\delta}{\sigma})^{-1/\xi} + \sum_{k=1}^{r_i}\{-\log\sigma-(1+1/\xi)\log(1+\xi\frac{z_i^{(k)}-\mu-\delta}{\sigma})\}\} \\
=\;& \sum_{i=1}^{m}\{-(1+\xi\frac{z_i^{(r_i)}-\mu}{\sigma})^{-1/\xi} + \sum_{k=1}^{r_i}\{-\log\sigma-(1+1/\xi)\log(1+\xi\frac{z_i^{(k)}-\mu}{\sigma})\}\} \\
+\;& \sum_{i=\tau+1}^{m}\{-(1+\xi\frac{z_i^{(r_i)}-\mu-\delta}{\sigma})^{-1/\xi} + (1+\xi\frac{z_i^{(r_i)}-\mu}{\sigma})^{-1/\xi} - (1+1/\xi)\sum_{k=1}^{r_i}\log\frac{1+\xi\frac{z_i^{(k)}-\mu-\delta}{\sigma}}{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}}\}
\end{aligned}
$$

Define $y_i = -(1+\xi\frac{z_i^{(r_i)}-\mu-\delta}{\sigma})^{-1/\xi} + (1+\xi\frac{z_i^{(r_i)}-\mu}{\sigma})^{-1/\xi} - (1+1/\xi)\sum_{k=1}^{r_i}\log\frac{1+\xi\frac{z_i^{(k)}-\mu-\delta}{\sigma}}{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}}$, and $\tilde{S}_k = \sum_{i=k+1}^{m} y_i$, we choose $\hat{\tau} = \arg\max_{k_0 \le k < m} \tilde{S}_k$. Therefore,

$$
\begin{aligned}
& \Lambda(\mu, \xi, \sigma) \\
= \;& \exp\{\sum_{i=\hat{\tau}+1}^{m} \{-(1+\xi\frac{z_i^{(r_i)}-\mu-\delta}{\sigma})^{-1/\xi} + (1+\xi\frac{z_i^{(r_i)}-\mu}{\sigma})^{-1/\xi} - (1+1/\xi)\sum_{k=1}^{r_i}\log\frac{1+\xi\frac{z_i^{(k)}-\mu-\delta}{\sigma}}{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}}\}\} \\
= \;& \exp\{\sum_{i=\hat{\tau}+1}^{m} y_i\} \\
= \;& \exp\{\max_{k_0 \le k < m} \tilde{S}_k\}
\end{aligned}
$$

Therefore, we reject when $\max_{k_0 \le k < m} \tilde{S}_k \ge L$ for some constant L.

To sum up, we reject either when $\exists j$, such that $\min(1+\xi\frac{z_j^{(1)}-\mu}{\sigma}, 1+\xi\frac{z_j^{(r_j)}-\mu}{\sigma}) \le 0$, or when $\forall j$, $\min(1+\xi\frac{z_j^{(1)}-\mu}{\sigma}, 1+\xi\frac{z_j^{(r_j)}-\mu}{\sigma}) > 0$, $\max_{k_0 \le k < m} \tilde{S}_k \ge L$.

When $\xi = 0$, the likelihood function for the null hypothesis is

$$
L_0(\mu, \sigma) = \prod_{i=1}^{m} \{\exp\{-\exp\{-\frac{z_i^{r_i}-\mu}{\sigma}\}\} \times \prod_{k=1}^{r_i} \sigma^{-1}\exp\{-\frac{z_i^{(k)}-\mu}{\sigma}\}\}
$$

The likelihood function for the alternative hypothesis is

$$
\begin{aligned}
L_1(\mu, \sigma) \;=\; & \prod_{i=1}^{\tau} \{\exp\{-\exp\{-\frac{z_i^{(r_i)}-\mu}{\sigma}\}\} \times \prod_{k=1}^{r_i} \sigma^{-1}\exp\{-\frac{z_i^{(k)}-\mu}{\sigma}\}\} \\
\times\; & \prod_{i=\tau+1}^{m} \{\exp\{-\exp\{-\frac{z_i^{(r_i)}-\mu-\delta}{\sigma}\}\} \times \prod_{k=1}^{r_i} \sigma^{-1}\exp\{-\frac{z_i^{(k)}-\mu-\delta}{\sigma}\}\}
\end{aligned}
$$

The likelihood ratio statistic is

$$
\Lambda(\mu, \sigma) = \frac{\max_{0 \le \tau < m} L_1(\tau, \mu, \sigma)}{\max_{\tau \ge m} L_0(\tau, \mu, \sigma)}
$$

So we need to maximize $L_1(\tau,\mu,\sigma)$.

$$\begin{aligned}
\log L_1(\tau,\mu,\sigma) &= \sum_{i=1}^{\tau}\{-\exp\{-\frac{z_i^{(r_i)}-\mu}{\sigma}\}+\sum_{k=1}^{r_i}\{-\log\sigma-\frac{z_i^{(k)}-\mu}{\sigma}\}\} \\
&+ \sum_{i=\tau+1}^{m}\{-\exp\{-\frac{z_i^{(r_i)}-\mu-\delta}{\sigma}\}+\sum_{k=1}^{r_i}\{-\log\sigma-\frac{z_i^{(k)}-\mu-\delta}{\sigma}\}\} \\
&= \sum_{i=1}^{m}\{-\exp\{-\frac{z_i^{(r_i)}-\mu}{\sigma}\}+\sum_{k=1}^{r_i}\{-\log\sigma-\frac{z_i^{(k)}-\mu}{\sigma}\}\} \\
&+ \sum_{i=\tau+1}^{m}\{\exp\{-\frac{z_i^{(r_i)}-\mu}{\sigma}\}-\exp\{-\frac{z_i^{(r_i)}-\mu-\delta}{\sigma}\}+\frac{r_i\delta}{\sigma}\}
\end{aligned}$$

Denote $y_i = \exp\{-\frac{z_i^{(r_i)}-\mu}{\sigma}\} - \exp\{-\frac{z_i^{(r_i)}-\mu-\delta}{\sigma}\} + \frac{r_i\delta}{\sigma}$, and $S_k = \sum_{i=1}^{k} y_i$, we choose $\hat{\tau} = arg\min_{0\le k<m} S_k$.

Therefore,

$$\Lambda(\hat{\tau},\mu,\sigma) = \sum_{i=\hat{\tau}+1}^{m} y_i = S_m - \min_{0\le k<m} S_k$$

We reject if $S_n - \min_{0\le k<m} S_k \ge L$ for some constant L.

For a change in all the parameters μ,σ and ξ, we separate the discussion into four scenarios: a) $\xi_1=0,\xi_2=0$. b) $\xi_1\neq 0,\xi_2\neq 0$. c) $\xi_1\neq 0,\xi_2=0$. d) $\xi_1=0,\xi_2\neq 0$.

$\underline{\boldsymbol{\xi_1\neq 0,\xi_2\neq 0}}$. The null hypothesis becomes

$$L_0(\mu,\xi,\sigma) = \prod_{i=1}^{m}\{\exp\{-\{1+\xi_1\frac{z_i^{(r_i)}-\mu_1}{\sigma_1}\}^{-1/\xi_1}\}\times\prod_{k=1}^{r_i}\sigma_1^{-1}\{1+\xi_1\frac{z_i^{(k)}-\mu_1}{\sigma_1}\}^{-1/\xi_1-1}\}$$

provided that $1+\xi_1\frac{z_i^{(k)}-\mu_1}{\sigma_1}>0$ for $k=1,2...,r_i$ and $i=1,2...,m$.

The alternative hypothesis is

$$\begin{aligned}
L_1(\mu,\xi,\sigma) &= \prod_{i=1}^{\tau}\{\exp\{-\{1+\xi_1\frac{z_i^{(r_i)}-\mu_1}{\sigma_1}\}^{-1/\xi_1}\}\times\prod_{k=1}^{r_i}\sigma_1^{-1}\{1+\xi_1\frac{z_i^{(k)}-\mu_1}{\sigma_1}\}^{-1/\xi_1-1}\} \\
&\times \prod_{i=\tau+1}^{m}\{\exp\{-\{1+\xi_2\frac{z_i^{(r_i)}-\mu_2}{\sigma_2}\}^{-1/\xi_2}\}\times\prod_{k=1}^{r_i}\sigma_2^{-1}\{1+\xi_2\frac{z_i^{(k)}-\mu_2}{\sigma_2}\}^{-1/\xi_2-1}\}
\end{aligned}$$

If $\exists j$, such that $\min(1+\xi_1\frac{z_j^{(1)}-\mu_1}{\sigma_1}, 1+\xi_1\frac{z_j^{(r_j)}-\mu_1}{\sigma_1}) \leq 0$, we reject the null because the likelihood for the null will be zero. Otherwise, we assume that $\forall j$, $\min(1+\xi_1\frac{z_j^{(1)}-\mu_1}{\sigma_1}, 1+\xi_1\frac{z_j^{(r_j)}-\mu_1}{\sigma_1}) > 0$ and $k_0 = \inf\{t : \min(1+\xi_2\frac{z_{t+1}^{(1)}-\mu_2}{\sigma_2}, 1+\xi_2\frac{z_{t+1}^{(r_{t+1})}-\mu_2}{\sigma_2}), \ldots, \min(1+\xi_2\frac{z_m^{(1)}-\mu_2}{\sigma_2}, 1+\xi_2\frac{z_m^{(r_m)}-\mu_2}{\sigma_2}) > 0\}$. Clearly, $\tau \geq k_0$.

The likelihood ratio statistic becomes

$$\begin{aligned}
& \Lambda(\mu,\xi,\sigma) = \frac{\max_{k_0\leq\tau<m} L_1(\tau)}{\max_{\tau\geq m} L_0(\tau)} \\
= \; & \{\prod_{i=1}^{m}\{\exp\{-\{1+\xi_1\frac{z_i^{(r_i)}-\mu_1}{\sigma_1}\}^{-1/\xi_1}\} \times \prod_{k=1}^{r_i}\sigma_1^{-1}\{1+\xi_1\frac{z_i^{(k)}-\mu_1}{\sigma_1}\}^{-1/\xi_1-1}\}\}^{-1} \\
\times \; & \prod_{i=1}^{\tau}\{\exp\{-\{1+\xi_1\frac{z_i^{(r_i)}-\mu_1}{\sigma_1}\}^{-1/\xi_1}\} \times \prod_{k=1}^{r_i}\sigma_1^{-1}\{1+\xi_1\frac{z_i^{(k)}-\mu_1}{\sigma_1}\}^{-1/\xi_1-1}\} \\
\times \; & \prod_{i=\tau+1}^{m}\{\exp\{-\{1+\xi_2\frac{z_i^{(r_i)}-\mu_2}{\sigma_2}\}^{-1/\xi_2}\} \times \prod_{k=1}^{r_i}\sigma_2^{-1}\{1+\xi_2\frac{z_i^{(k)}-\mu_2}{\sigma_2}\}^{-1/\xi_2-1}\} \\
\propto \; & \prod_{i=1}^{\tau}\{\exp\{-\{1+\xi_1\frac{z_i^{(r_i)}-\mu_1}{\sigma_1}\}^{-1/\xi_1}\} \times \prod_{k=1}^{r_i}\sigma_1^{-1}\{1+\xi_1\frac{z_i^{(k)}-\mu_1}{\sigma_1}\}^{-1/\xi_1-1}\} \\
\times \; & \prod_{i=\tau+1}^{m}\{\exp\{-\{1+\xi_2\frac{z_i^{(r_i)}-\mu_2}{\sigma_2}\}^{-1/\xi_2}\} \times \prod_{k=1}^{r_i}\sigma_2^{-1}\{1+\xi_2\frac{z_i^{(k)}-\mu_2}{\sigma_2}\}^{-1/\xi_2-1}\}
\end{aligned}$$

Maximizing Λ is equivalent to maximizing $\log\Lambda$.

$$\begin{aligned}
\log\Lambda \; = \; & \sum_{i=1}^{\tau}\{-(1+\xi_1\frac{z_i^{(r_i)}-\mu_1}{\sigma_1})^{-1/\xi_1} + \sum_{k=1}^{r_i}\{-\log\sigma_1 - (1+1/\xi_1)\log(1+\xi_1\frac{z_i^{(k)}-\mu_1}{\sigma_1})\}\} \\
+ \; & \sum_{i=\tau+1}^{m}\{-(1+\xi_2\frac{z_i^{(k)}-\mu_2}{\sigma_2})^{-1/\xi_2} + \sum_{k=1}^{r_i}\{-\log\sigma_2 - (1+1/\xi_2)\log(1+\xi_2\frac{z_i^{(k)}-\mu_2}{\sigma_2})\}\} \\
= \; & \sum_{i=1}^{m}\{-(1+\xi_1\frac{z_i^{(r_i)}-\mu_1}{\sigma_1})^{-1/\xi_1} + \sum_{k=1}^{r_i}\{-\log\sigma_1 - (1+1/\xi_1)\log(1+\xi_1\frac{z_i^{(k)}-\mu_1}{\sigma_1})\}\} \\
+ \; & \sum_{i=\tau+1}^{m}\{-(1+\xi_2\frac{z_i^{(r_i)}-\mu_2}{\sigma_2})^{-1/\xi_2} + (1+\xi_1\frac{z_i^{(r_i)}-\mu_1}{\sigma_1})^{-1/\xi_1} \\
+ \; & \sum_{k=1}^{r_i}\{\log\sigma_1 + (1+\frac{1}{\xi_1})\log(1+\xi_1\frac{z_i^{(k)}-\mu_1}{\sigma_1}) - \log\sigma_2 - (1+\frac{1}{\xi_2})\log(1+\xi_2\frac{z_i^{(k)}-\mu_2}{\sigma_2})\}\}
\end{aligned}$$

Define

$$y_i = -(1+\xi_2\frac{z_i^{(r_i)}-\mu_2}{\sigma_2})^{-1/\xi_2} + (1+\xi_1\frac{z_i^{(r_i)}-\mu_1}{\sigma_1})^{-1/\xi_1}$$
$$+ \quad \sum_{k=1}^{r_i}\{\log\sigma_1 + (1+\frac{1}{\xi_1})\log(1+\xi_1\frac{z_i^{(k)}-\mu_1}{\sigma_1}) - \log\sigma_2 - (1+\frac{1}{\xi_2})\log(1+\xi_2\frac{z_i^{(k)}-\mu_2}{\sigma_2})\}$$

and $\tilde{S}_k = \sum_{i=k+1}^{m} y_i$, we choose $\hat{\tau} = arg\max_{k_0\le k<m}\tilde{S}_k$.

It is easy to show that

$$\Lambda(\hat{\tau}) = \exp\{\tilde{S}_{\hat{\tau}}\} = \exp\{\max_{k_0\le k<m}\tilde{S}_k\}$$

We reject when $\max_{k_0\le k<m}\tilde{S}_k \ge L$.

To sum up, If $\exists j$, such that $\min(1+\xi_1\frac{z_j^{(1)}-\mu_1}{\sigma_1}, 1+\xi_1\frac{z_j^{(r_j)}-\mu_1}{\sigma_1}) \le 0$, we reject the null, otherwise, we reject if $\max_{k_0\le k<m}\tilde{S}_k$ exceeds some constant L.

$\underline{\boldsymbol{\xi_1 = 0, \xi_2 = 0}}$. The likelihood under the null hypothesis is

$$L_0(\mu,\sigma) = \prod_{i=1}^{m}\{\exp\{-\exp\{-\frac{z_i^{r_i}-\mu_1}{\sigma_1}\}\} \times \prod_{k=1}^{r_i}\sigma_1^{-1}\exp\{-\frac{z_i^{(k)}-\mu_1}{\sigma_1}\}\}$$

The likelihood for the alternative is

$$\begin{aligned} L_1(\mu_1,\sigma_1,\mu_2,\sigma_2) \quad &= \quad \prod_{i=1}^{\tau}\{\exp\{-\exp\{-\frac{z_i^{(r_i)}-\mu_1}{\sigma_1}\}\} \times \prod_{k=1}^{r_i}\sigma_1^{-1}\exp\{-\frac{z_i^{(k)}-\mu_1}{\sigma_1}\}\} \\ &\times \prod_{i=\tau+1}^{m}\{\exp\{-\exp\{-\frac{z_i^{(r_i)}-\mu_2}{\sigma_2}\}\} \times \prod_{k=1}^{r_i}\sigma_2^{-1}\exp\{-\frac{z_i^{(k)}-\mu_2}{\sigma_2}\}\} \end{aligned}$$

The likelihood ratio statistic is

$$\Lambda(\mu_1,\sigma_1,\mu_2,\sigma_2) = \frac{\max_{0\le\tau<m} L_1(\tau,\mu_1,\sigma_1,\mu_2,\sigma_2)}{\max_{\tau\ge m} L_0(\tau,\mu_1,\sigma_1)} \propto \max_{0\le\tau<m} L_1(\tau,\mu_1,\sigma_1,\mu_2,\sigma_2)$$

So we need to maximize $L_1(\tau, \mu_1, \sigma_1, \mu_2, \sigma_2)$.

$$
\begin{aligned}
& \log L_1(\tau, \mu_1, \sigma_1, \mu_2, \sigma_2) \\
= & \sum_{i=1}^{\tau}\{-\exp\{-\frac{z_i^{(r_i)}-\mu_1}{\sigma_1}\} + \sum_{k=1}^{r_i}\{-\log\sigma_1 - \frac{z_i^{(k)}-\mu_1}{\sigma_1}\}\} \\
+ & \sum_{i=\tau+1}^{m}\{-\exp\{-\frac{z_i^{(r_i)}-\mu_2}{\sigma_2}\} + \sum_{k=1}^{r_i}\{-\log\sigma_2 - \frac{z_i^{(k)}-\mu_2}{\sigma_2}\}\} \\
= & \sum_{i=1}^{m}\{-\exp\{-\frac{z_i^{(r_i)}-\mu_1}{\sigma_1}\} + \sum_{k=1}^{r_i}\{-\log\sigma_1 - \frac{z_i^{(k)}-\mu_1}{\sigma_1}\}\} \\
+ & \sum_{i=\tau+1}^{m}\{\exp\{-\frac{z_i^{(r_i)}-\mu_1}{\sigma_1}\} - \exp\{-\frac{z_i^{(r_i)}-\mu_2}{\sigma_2}\} + \sum_{k=1}^{r_i}\{\log\sigma_1 - \log\sigma_2 + \frac{z_i^{(k)}-\mu_1}{\sigma_1} - \frac{z_i^{(k)}-\mu_2}{\sigma_2}\}
\end{aligned}
$$

Denote $y_i = \exp\{-\frac{z_i^{(r_i)}-\mu_1}{\sigma_1}\} - \exp\{-\frac{z_i^{(r_i)}-\mu_2}{\sigma_2}\} + \sum_{k=1}^{r_i}\{\log\sigma_1 - \log\sigma_2 + \frac{z_i^{(k)}-\mu_1}{\sigma_1} - \frac{z_i^{(k)}-\mu_2}{\sigma_2}\}$, and $\tilde{S}_k = \sum_{i=k+1}^{m} y_i$, we choose $\hat{\tau} = arg\max_{0\le k<m} \tilde{S}_k$.

Therefore,

$$
\Lambda(\hat{\tau}, \mu_1, \sigma_1, \mu_2, \sigma_2) = \sum_{i=\hat{\tau}+1}^{m} y_i = \max_{0\le k<m} \tilde{S}_k
$$

We reject if $\max_{0\le k<m} \tilde{S}_k \ge L$ for some constant L.

$\underline{\boldsymbol{\xi_1 = 0, \xi_2 \neq 0}}$. Define

$$
\begin{aligned}
& y_i = -(1+\xi_2\frac{z_i^{(r_i)}-\mu_2}{\sigma_2})^{-1/\xi_2} + \exp\{-\frac{z_i^{(r_i)}-\mu_1}{\sigma_1}\} \\
+ & \sum_{k=1}^{r_i}\{\log\sigma_1 + \frac{z_i^{(k)}-\mu_1}{\sigma_1} - \log\sigma_2 - (1+\frac{1}{\xi_2})\log(1+\xi_2\frac{z_i^{(k)}-\mu_2}{\sigma_2})\}
\end{aligned}
$$

Define $k_0 = \inf\{t : 1+\xi_2\frac{x_{t+1}-\mu_2}{\sigma_2} > 0, 1+\xi_2\frac{x_{t+2}-\mu_2}{\sigma_2} > 0, \ldots, 1+\xi_2\frac{x_n-\mu_2}{\sigma_2} > 0\}$, and $\tilde{S}_k = \sum_{i=k+1}^{n} y_i$, we reject if $\max_{k_0\le k<n} \tilde{S}_k \ge L$.

$\underline{\boldsymbol{\xi_1 \neq 0, \xi_2 = 0}}$. Define

$$y_i = -\exp\{-\frac{z_i^{(r_i)} - \mu_2}{\sigma_2}\} + (1 + \xi_1 \frac{z_i^{(r_i)} - \mu_1}{\sigma_1})^{-1/\xi_1}$$
$$+ \sum_{k=1}^{r_i} \{\log \sigma_1 + (1 + \frac{1}{\xi_1}) \log(1 + \xi_1 \frac{z_i^{(k)} - \mu_1}{\sigma_1}) - \log \sigma_2 - \frac{z_i^{(k)} - \mu_2}{\sigma_2}\}$$

The conclusion being if $\exists j$, such that $1 + \xi_1 \frac{x_j - \mu_1}{\sigma_1} \leq 0$, we reject the null hypothesis. Otherwise, define $\tilde{S}_k = \sum_{i=k+1}^{n} y_i$, we reject when $\max_{0 \leq k < n} \tilde{S}_k \geq L$.

4.3 Generating the R-order Statistic

In order to perform inference on change detection using order statistic, we need to compute the p-value to indicate the strength of evidence for a change. Since p-value is computed under the null case, i.e. when there is no change, it is important to simulate order statistic by blocks under the null scenario.

To formulate mathematically, given the number of order statistic r, our objective is to simulate $(z_i^{(1)}, z_i^{(2)}, \ldots, z_i^{(r)})$ within each block i, where $z^{(1)} \geq z^{(2)} \ldots \geq z^{(r)}$ and $(z_i^{(1)}, z_i^{(2)}, \ldots, z_i^{(r)})$ are the maximum r-order statistic. The joint probability function for $(z_i^{(1)}, z_i^{(2)}, \ldots, z_i^{(r)})$ is given by:

$$f(z^{(1)}, z^{(2)}, \ldots, z^{(r)}) = f(z^{(1)}) f(z^{(2)}|z^{(1)}) \ldots f(z^{(r)}|z^{(1)}, \ldots, z^{(r-1)}) \quad (4.3.5)$$

Inspired by Equation 4.3.5, a sequential generating algorithm can be designed using rejection sampling. Conditional on $z_i^{(1)}, \ldots, z_i^{(k-1)}$, we generate $z_i^{(k)}$ for $k = 2, \ldots, r$. The following algorithm is proposed:

Step 0: Simulate $z_i^{(1)}$ by generating a random number from $gev(\mu, \sigma, \xi)$. This is due to the fact that $z_i^{(1)}$ is the maxima within block i. R packages *evd* has the function *rgev* to perform this task.

for $j = 2, \ldots, r$:

Step 1: compute $f(z_i^{(j)}|z_i^{(1)}, \ldots z_i^{(j-1)})$.

By [22], the joint probability density function for $(z_i^{(1)}, \ldots, z_i^{(r)})$ is

$$
\begin{aligned}
& f(z_i^{(1)}, ...z_i^{(r-1)}, z_i^{(r)}) \\
= \ & \exp\{-\{1+\xi\frac{z_i^{(r)}-\mu}{\sigma}\}^{-1/\xi}\}\prod_{k=1}^{r}\{\sigma^{-1}\{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}\}^{-1-1/\xi}I_{\{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}>0\}}\}I_{\{z_i^{(r)}\le\ldots\le z_i^{(1)}\}}
\end{aligned}
$$

Therefore, the conditional probability of $f(z_i^{(j)}|z_i^{(1)}, ...z_i^{(j-1)})$ is given by

$$
\begin{aligned}
& f(z_i^{(j)}|z_i^{(1)}, ...z_i^{(j-1)}) = f(z_i^{(1)}, ...z_i^{(j-1)})^{-1} f(z_i^{(1)}, ...z_i^{(j-1)}, z_i^{(j)}) \\
= \ & \{\exp\{-\{1+\xi\frac{z_i^{(j-1)}-\mu}{\sigma}\}^{-1/\xi}\}\prod_{k=1}^{j-1}\{\sigma^{-1}\{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}\}^{-1-1/\xi}I_{\{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}>0\}}\}I_{\{z_i^{(j-1)}\le\ldots\le z_i^{(1)}\}}\}^{-1} \\
\times \ & \exp\{-\{1+\xi\frac{z_i^{(j)}-\mu}{\sigma}\}^{-1/\xi}\}\prod_{k=1}^{j}\{\sigma^{-1}\{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}\}^{-1-1/\xi}I_{\{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}>0\}}\}I_{\{z_i^{(j)}\le\ldots\le z_i^{(1)}\}} \\
= \ & \exp\{-\{1+\xi\frac{z_i^{(j)}-\mu}{\sigma}\}^{-1/\xi}+\{1+\xi\frac{z_i^{(j-1)}-\mu}{\sigma}\}^{-1/\xi}\} \\
\times \ & \sigma^{-1}\{1+\xi\frac{z_i^{(j)-\mu}}{\sigma}\}^{-1-1/\xi}I_{\{1+\xi\frac{z_i^{(j)}-\mu}{\sigma}>0\}}I_{\{z_i^{(j)}\le z_i^{(j-1)}\}}
\end{aligned}
$$

Think of it as $f(x|z_i^{(1)}, ...z_i^{(j-1)})$.

Step 2: Take

$$g(x) = \exp\{-\{1+\xi\frac{x-\mu}{\sigma}\}^{-1/\xi}\}\sigma^{-1}\{1+\xi\frac{x-\mu}{\sigma}\}^{-1-1/\xi}I_{\{1+\xi\frac{x-\mu}{\sigma}>0\}}$$

In the rejection sampling, it is necessary to specify another function $g(x)$, from which it is easy to draw samples. Here we choose $g(x)$ as the probability density function of $gev(\mu, \sigma, \xi)$.

Step 3: Draw a sample x from $g(x)$, which is $gev(\mu, \sigma, \xi)$. This step is identical as Step 0.

Step 4: Determine the constant c

$$c = \exp\{\{1+\xi\frac{z_i^{(i-1)}-\mu}{\sigma}\}^{-1/\xi}\}$$

Step 5: Compute the probability of including x in our sample pool:

$$r = \frac{f(x|z_i^{(i-1)}, \ldots, z_i^{(1)})}{cg(x)} = I_{\{x\le z_i^{(i-1)}\}}$$

It turns out that r is an indicator function having values of either 0 or 1.

Step 6: If $x \leq z_i^{(i-1)}$, we accept x with probability 1. Otherwise, we reject x with probability 1 and go back to Step 3.

end

In summary, for each block $1 \leq i \leq m$, we may simulate $(z_i^{(1)}, \ldots, z_i^{(r)})$ and save them for later analysis.

4.4 Maximum Likelihood Estimation

This section is devoted to computing the maximum likelihood estimation. As stated in Equation 4.1.1, the joint distribution for $(z^{(1)}, \ldots, z^{(r)})$ is

$$
\begin{aligned}
f(z^{(1)}, \ldots, z^{(r)}) &= \exp\{-\{1+\xi\frac{z^{(r)}-\mu}{\sigma}\}^{-1/\xi}\} \\
&\times \prod_{k=1}^{r} \sigma^{-1}\{1+\xi\frac{z^{(k)}-\mu}{\sigma}\}^{-1-1/\xi}
\end{aligned}
$$

given $1+\xi\frac{z^{(k)}-\mu}{\sigma} > 0$ for $k = 1, 2, \ldots, r$, and $\sigma > 0$.

Therefore, the likelihood for the observed data is

$$
\begin{aligned}
L(z_i^{(1)}, \ldots, z_i^{(r)}, i = 1, 2, \ldots, m) &= \prod_{i=1}^{m} \exp\{-\{1+\xi\frac{z_i^{(r)}-\mu}{\sigma}\}^{-1/\xi}\} \\
&\times \prod_{k=1}^{r} \sigma^{-1}\{1+\xi\frac{z_i^{(k)}-\mu}{\sigma}\}^{-1-1/\xi}
\end{aligned}
$$

given $1+\xi\frac{z_i^{(k)}-\mu}{\sigma} > 0$ for $k = 1, 2, \ldots, r$, $i = 1, 2, \ldots, m$, and $\sigma > 0$.

To compute the maximum likelihood $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ estimator, we should maximize L under the constraint that $1+\xi\frac{z_i^{(k)}-\mu}{\sigma} > 0$ for $k = 1, 2, \ldots, r$, and $i = 1, 2, \ldots, m$. This is a constraint optimization problem and the constraint is non linear. However, we manage to transform the above problem into a constraint optimization with linear constraints by transforming the parameters.

$$\begin{cases} \theta_1 = \xi/\sigma \\ \theta_2 = \xi\mu/\sigma \\ \theta_3 = \sigma \end{cases} \tag{4.4.6}$$

The constraints become:

$$\begin{cases} 1 + \theta_1 z_i^{(k)} - \theta_2 > 0 \;\; \forall i = 1, \ldots, m; k = 1, 2, \ldots, r. \\ \theta_3 > 0 \end{cases}$$

Written in the matrix form, it becomes $U\theta - c \geq 0$, where $U = \begin{pmatrix} z_1^{(1)} & -1 & 0 \\ z_1^{(2)} & -1 & 0 \\ \ldots & & \\ z_i^{(r)} & -1 & 0 \\ \ldots & & \\ z_m^{(1)} & -1 & 0 \\ .. & & \\ z_m^{(r)} & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Put $c = (-1+\epsilon, \ldots, -1+\epsilon, \epsilon)'$, where ϵ is a very small positive number. Here we use $\epsilon = 10^{-6}$.

The objective function becomes

$$L = \prod_{i=1}^{m} \exp\{-\{1 + \theta_1 z_i^{(r)} - \theta_2\}^{-1/(\theta_1\theta_3)}\} \prod_{k=1}^{r} \theta_3^{-1}\{1 + \theta_1 z_i^{(k)} - \theta_2\}^{-1-1/(\theta 1 \theta 3)}$$

Now we implement the *constrOptim* function in the R package, and solve $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$. After this step, according to Equation 4.4.6, we transform those estimated parameters backwards and solve for the original parameters $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$.

$$\begin{cases} \hat{\mu} = \hat{\theta}_2/\hat{\theta}_1 \\ \hat{\sigma} = \hat{\theta}_3 \\ \hat{\xi} = \hat{\theta}_1\hat{\theta}_3 \end{cases}$$

4.5 Hurricane Data Analysis with Order Statistic

This section revisits the change detection problem on maximum sustained wind speeds with the procedure described above.

One change point model is assumed. The data set consists of r-largest sustained wind speeds across all Atlantic hurricanes each year from 1951-2008. Year is considered blocks in this example, and we take year 1951 to be block 1, ..., year 2008 to be block 58. Therefore, the data set consists of 58 blocks, and within each block $1 \leq i \leq 58$, there are r observations, which are the r largest sustained wind speeds across all hurricanes for a year.

To formulate mathematically, we observe $(z_i^{(1)}, \ldots, z_i^{(r)})$, where $i = 1, 2, \ldots, T = 58$. We consider $r = 1, 2, ..5$ respectively for our analysis. Note that in the case of $r = 1$, we go back to the original single maxima case.

The joint density function of $(z_i^{(1)}, \ldots, z_i^{(r)})$ is given by Equation 4.1.1 and our task is to detect whether there is any change in the vector (μ, σ, ξ). Our proposed change detection methodology is straightforward: under the assumption that there is at most one change point in the data series, if there is actually a change in the data series at time point t, where $1 \leq t \leq T$, the parameter estimates before time point and after t should be as separated as possible. To measure the distance of the two parameter estimates, we employ the popular L_2 norm, and compute $d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$, where n is the dimension of the x and y vector. In this example, $n = 3$, and $x = (\mu_1, \sigma_1, \xi_1)$ and $y = (\mu_2, \sigma_2, \xi_2)$. Based on the real data, we choose $\hat{t}$ that best separates the two phases.

Another point in our proposed methodology is to quantify how large in the difference is considered significant. Due to the randomness of the data, we need to understand whether the computed difference from the real data is large enough to conclude statistical significance of a change in the data. To do this, we simulate data from the no change model using parametric bootstrap, and in each simulation, we compute the maximum difference. A p-value is thus obtained for the standardized maximum difference computed from the real data. A small p-value indicates strong evidence of a change, while a large p-value indicates we retain the idea of no change.

Here is the algorithm:

Step 1: Select the appropriate $t_0 = 10$, $t_1 = 50$. This serves as a penalty that the change point t can only occur between t_0 and t_1. Although this might not be the case in the real data, this constraint is necessary in that if t is too close to 1 or T, our parameter estimation is highly biased and inaccurate.

Step 2: For t=$t_0, \ldots,\ t_1$, perform maximum likelihood estimation on the data series $1, 2, \ldots, t$ and $t+1, \ldots, T$. This gives us $\hat{\theta}_0^t$ and $\hat{\theta}_1^t$. Compute $D(t) = ||\hat{\theta}_0^t - \hat{\theta}_1^t||$, where $||x||$ is denoted as the Euclidean distance.

Step 3: Find the t which maximizes $D(t)$. That is the time point that best separates the data series into two phases. Here $t_0 \leq t \leq t_1$. We denote the estimate as $\hat{t}$, and record the maximum likelihood estimator for data $1, 2, \ldots, \hat{t}$ as $\theta_0^{\hat{t}}$, and the maximum likelihood estimator for data $1, 2, \ldots, \hat{t}$ as $\theta_1^{\hat{t}}$. Compute $\delta = \theta_0^{\hat{t}} - \theta_1^{\hat{t}}$.

Step 4: Choose $k_0 = 0$ and $k_0 = 5$ for generating the data series under the null case when no change occurs. k_0 gives us room for errors in maximum likelihood estimation. To be more specific, if $\hat{t}$ is greater than $(t_0 + t_1)/2$, we generate data from the maximum likelihood estimation of the series $1, 2, \ldots, \hat{t} - k_0$. If $\hat{t}$ is equal to or smaller than $(t_0 + t_1)/2$, we generate data from the maximum likelihood estimation of the series $\hat{t} + k_0, \ldots, T$. Denote the maximum likelihood estimator to be θ_0^*.

Step 5: In Step 4, we generate data using the rejection sampling methodology described in the previous section with the parameter of θ_0^*. We simulate B times and get B sets from data.

Step 6: For each generated data $1 \leq b \leq B$, redo Step 2-4, and record $\delta_{(b)} = \theta_0^{\hat{t}_b} - \theta_1^{\hat{t}_b}$.

Step 7: Based $\delta_{(b)}$, where $b = 1, 2, \ldots, 100$, compute the variance covariance matrix, which we denote as Σ.

Step 8: Compute $c = \delta^\top \Sigma^{-1} \delta$, and $c_{(b)} = \delta_{(b)}^\top \Sigma^{-1} \delta_{(b)}$.

Step 9: Compute the quantile of c against $c_{(b)}$, and return p-value.

The results are summarized in Table 4.1, 4.2 and 4.3. In the table, r is the number of order statistic to consider, $\hat{t}$ denotes the time point that maximizes the difference of the two *gev* parameter estimates. Test statistic represents the value of c, and p-value is the tail probability of c against $c_{(b)}$ for $b = 1, 2, \ldots, 100$. It is clearly seen that as r increases, the p-value in general decreases. This is consistent with the statement we made: the more order statistics is involved in the analysis, the more effective the testing procedure is. It is clearly shown that there is a change in the wind speed, and by leveraging order statistic in the change detection methodology, the evidence gets stronger.

It is also worth noting that $\hat{t}$ is very stable except for $r = 5$. For $r = 1, 2, 3, 4$, $\hat{t}$ is between 10 and 20. This gives consistent result that there is a change in the wind speed during the 1960s and 1970s. For the exceptionally large value of $\hat{t} = 47$ when $r = 5$, it is most likely that the GEV distribution is not a good approximation to the real data when we consider five maximum order statistics together. Therefore, we might drop the case for $r = 5$.

r value	$\hat{t}$	$\hat{\theta}_0$	$\hat{\theta}_1$
$r = 1$	12	(139.69,21.03,-2.04)	(118.74,21.86, -0.36)
$r = 2$	14	(139.92,8.70,-0.84)	(129.52,17.83, -0.35)
$r = 3$	19	(140.47,13.52,-0.47)	(129.63,18.39,-0.04)
$r = 4$	11	(139.25,7.27,-0.58)	(135.94,19.61, -0.11)
$r = 5$	47	(136.90,13.63,-0.39)	(147.89,17.19,0.049)

Table 4.1: Estimation for parameters using order statistic in the study of maximum sustained wind speeds. r is the number of order statistics to consider, $\hat{t}$ is the estimation of change point by using the maximum separation criteria, $\hat{\theta}_0$ and $\hat{\theta}_1$ are the parameter estimations before and after the estimated change point $\hat{t}$ respectively. $t_0 = 10$, $t_1 = 50$.

r value	$\hat{t}$	Test Statistic c	p-value
$r = 1$	12	4.850246	0.13
$r = 2$	14	10.09187	0.03
$r = 3$	19	9.12865	0.05
$r = 4$	11	29.73115	0
$r = 5$	47	28.15609	0

Table 4.2: P-value approach for sustained wind speeds: We use $t_0 = 10, t_1 = 50, k_0 = 0$ and conduct 100 simulations in generating the null data.

r value	$\hat{t}$	Test Statistic c	p-value
$r = 1$	12	5.832056	0.12
$r = 2$	14	16.55036	0.02
$r = 3$	19	5.651442	0.09
$r = 4$	11	12.04822	0
$r = 5$	47	34.64381	0

Table 4.3: P-value Approach for sustained wind speeds: We use $t_0 = 10, t_1 = 50, k_0 = 5$ and conduct 100 simulations in generating the null data.

Chapter 5

Hurricane Trajectory Prediction

This chapter leverages machine learning approaches for the hurricane trajectory prediction. Here we address two variants of this problem: m step ahead sequential prediction and m step ahead direct prediction, where m is arbitrary, but typically 4, 8 and so on. In Section 5.1, we present the background of this problem, data preparation and introduction to the collaborative filtering approach as one of our major learning procedures. In Section 5.2, we present seven individual algorithms for predicting the next location of the target hurricane, with each algorithm using either supervised, unsupervised or hybrid approach. In Section 5.3, we discuss fifteen weighting schemes for combining those prediction results, with some of them dynamically updating the weight of each algorithm based on their past performance. In Section 5.4, we discuss m step ahead direct prediction procedure. In Section 5.5, we present the performance for the fifteen weighting schemes, with the discussion of some of the advantages and disadvantages. The 2nd hurricane in 1952 and 4th in 2008 serve as two examples to illustrate our algorithmic performance under different scenarios. Section 5.6 presents future research directions.

5.1 Problem Formulation

Before analysis, we perform the transformation of the raw latitude and longitude to the Cartesian coordinates due to several drawbacks of raw longitude and latitude.

$$XCart = cos(latitude\ \pi/180)cos(longitude\ \pi/180) \tag{5.1.1}$$

$$YCart = cos(latitude\,\pi/180)sin(longitude\,\pi/180) \tag{5.1.2}$$

The drawbacks of using the raw latitude and longitude for analysis are three fold: Firstly, they are measured in degrees, and five degrees of longitude difference near the equator is much longer than five degrees of longitude difference near the north pole, making latitude and longitude an inaccurate measure of distance traveled. Cartesian transformation helps mitigate this problem. Secondly, for longitude, the degree unit has a period of 360, and Cartesian transformation with trigonometric function built in is perfect for this purpose. Thirdly, Cartesian transformation essentially normalizes the raw latitude and longitude to [-1,1]. We will be using Cartesian coordinates throughout this chapter, and refer to latitude and longitude as convention.

We propose both the supervised and unsupervised approach to build m-step ahead sequential prediction. Specifically, for the unsupervised approach, we employ collaborative filtering because we would like to leverage past hurricanes' trajectories. If we find a set of past hurricanes whose trajectories behave similarly to the target hurricane for the initial k points, we may reasonably argue that the target hurricane might behave more or less the same way in the future as the trajectory of the past hurricanes.

Assume k is any fixed integer number. For the two hurricanes with (longitude, latitude) pair at time point $t = 1, 2, \ldots, k$, which we denote as $\{X^1(t), Y^1(t)\}$ and $\{X^2(t), Y^2(t)\}$, we define the first difference to be: $X_1^1(t) = X^1(t) - X^1(t-1)$, $Y_1^1(t) = Y^1(t) - Y^1(t-1)$, $X_1^2(t) = X^2(t) - X^2(t-1)$, $Y_1^2(t) = Y^2(t) - Y^2(t-1)$. Those quantities essentially represent the hurricane velocity at time t in the x-axis and y-axis.

We define the *direction* of a hurricane i to be $(V^i(t), W^i(t))$ at time point t:

If at some time point t, $X_1^i(t) = 0$ and $Y_1^i(t) = 0$ in which hurricane stay still, define $(V^i(t), W^i(t)) = (0, 0)$. Otherwise define

$$(V^i(t), W^i(t)) = (\frac{X_1^i(t)}{\sqrt{X_1^i(t)^2 + Y_1^i(t)^2}}, \frac{Y_1^i(t)}{\sqrt{X_1^i(t)^2 + Y_1^i(t)^2}}) \tag{5.1.3}$$

And for hurricane i and j,

$$Sim((X^i, Y^i), (X^j, Y^j)) = Sim((V^i, W^i), (V^j, W^j)) = \sum_{t=2}^{k} (V^i(t) - V^j(t))^2 + (W^i(t) - W^j(t))^2 \tag{5.1.4}$$

The similarity between the two hurricanes from our perspective is defined as the closeness of the directions at each time point from $t = 1$ to $t = k$. From Equation 5.1.4, the smaller the similarity measure, the closer the direction between the two. Note that this is one interpretation of similar hurricanes, and the idea can be extended to include lags. For example, hurricane j might not behave similarly as hurricane i at time point $t = 1$, but starting from $t = 2$ and onwards, it behaves similarly as hurricane i starting from $t = 1$. Incorporating the lagging effect into the similarity measure will include more quality hurricanes into the candidate set, thus improving prediction. We will discuss this in later sections.

After defining similarity, we examine all the past hurricanes j prior to the target hurricane i, and assess the similarity between each of them and the target one. The similarity measure can be transformed into a weight measure, which we define as

$$weight_j = e^{-Sim((X^i,Y^i),(X^j,Y^j))} \tag{5.1.5}$$

For the sake of simplicity and noise removal, we select only the top fifty percent of the past hurricanes in terms of weights into our candidate set J_1. For each candidate hurricane, we put together their past speed information, their next location/speed/direction as well as their normalized weight into a weighted average algorithm to estimate the location of the next movement for the target hurricane. We will show some of the algorithms below to reflect this philosophy.

Another perspective for defining the hurricane similarity is based on the speed. We define the *speed* of a hurricane i to be $S^i(t)$ at time point t:

If for some time point t, $X_1^i(t) = 0$ and $Y_1^i(t) = 0$ in which hurricane stay still, define $S^i(t) = 0$. Otherwise define

$$S^i(t) = \sqrt{X_1^i(t)^2 + Y_1^i(t)^2} \tag{5.1.6}$$

And for hurricane i and j,

$$Sim(S^i, S^j) = corr(S^i, S^j) \tag{5.1.7}$$

And

$$weight_j = max(Sim(S^i, S^j), 0) \tag{5.1.8}$$

Like the previous approach, we can select top fifty percent of the most similar hurricanes to be in the candidate set J_2.

Aside from collaborative filtering, we also employ supervised learning approach based on the target hurricane's past trajectory. Specifically, by building a relationship between its speed and its acceleration, we aim to predict the speed in its next movement. Numerically, the speed (with direction) is the first difference we have defined before: $(X_1^i(t), Y_1^i(t))$, and acceleration is defined as the second difference: $X_2^i(t) = X_1^i(t) - X_1^i(t-1)$, $Y_2^i(t) = Y_1^i(t) - Y_1^i(t-1)$. The Markov property is assumed: speed at time t only depends on the speed at $t-1$, and acceleration at $t-1$, which in turn involves speed at time $t-1$ and $t-2$. From another perspective, we are essentially constructing an AR(2) time series model for speed. On the other hand, in the regression context, we may model $X_1^i(t)$ on $X_1^i(t-1) + X_2^i(t-1)$. Of course, an even simpler approach is to assume the hurricane to be *momentum invariant*: speed at time t is equal to the speed at $t-1$, or some AR(1) model, which we also include in one of the algorithms. This might hold true for some regular hurricanes and for the purpose of short term prediction, but as we get more irregular shaped hurricanes or when we want to predict a few steps ahead, this may not work as effectively. In some of the algorithms we propose below, we also implement some hybrid approach to include both supervised and unsupervised learning.

Finally, we define the error measure, which we propose either the absolute error or the mean squared error:

$$e_i = \sum_{t=k}^{k+m-1} (\hat{X}^i(t+1) - X^i(t+1))^2 \tag{5.1.9}$$

Or

$$e_i = \sum_{t=k}^{k+m-1} |\hat{X}^i(t+1) - X^i(t+1)| \tag{5.1.10}$$

5.2 Algorithms

We will show seven algorithms that perform individual m step ahead sequential prediction for hurricane i, assuming we observe $t = 1, \ldots, k$. Since m step ahead sequential prediction is essentially repeating one step ahead prediction for m times with adjustment, we focus on one step ahead prediction problem with the following algorithms.

Algorithm One:

Assumption: the speed ratio at time point t will remain at $t+1$ between all past hurricanes j and hurricane i, and direction is used for similarity.

For each hurricane j

For each time $t = k, \ldots, k+m-1$

Compute the ratio of the mean speed of past hurricane j in the candidate set J_1 and the mean speed of the current hurricane i until time t

Compute the prediction of hurricane i at time point $t+1$ to be:

$$\hat{X}^i(t+1) = X^i(t) + sum_{j \in J_1} weight_j(direction) speedratio_j X_1^j(t+1) \tag{5.2.11}$$

$$\hat{Y}^i(t+1) = Y^i(t) + sum_{j \in J_1} weight_j(direction) speedratio_j Y_1^j(t+1) \tag{5.2.12}$$

Algorithm Two:

Assumption: The linear relationship between speed of hurricane j and that of hurricane i until time t stays the same from t to $t+1$, with recent data points bearing more weight, and direction is used for similarity.

For each hurricane j

Compute weighted least square regression of $S^i(t)$ on $S^j(t)$ for $t = 2, \ldots, k$, with weights $e^{-(k-t+1)}$. Denote intercept to be α_j and β_j.

Compute prediction of hurricane i at time point $t+1$ to be:

$$\hat{X}^i(t+1) = X^i(t) + sum_{j \in J_1} weight_j(direction)(\alpha_j + \beta_j S^j(t+1)) V^j(t+1)/S^j(t+1) \tag{5.2.13}$$

$$\hat{Y}^i(t+1) = Y^i(t) + sum_{j \in J_1} weight_j(direction)(\alpha_j + \beta_j S^j(t+1)) W^j(t+1)/S^j(t+1) \tag{5.2.14}$$

Algorithm Three:

Assumption: the speed ratio at time point t will remain at $t+1$ for all past hurricanes j and speed is used for similarity.

Similar to Algorithm One, with changes from direction based similarity to speed based similarity:

$$\hat{X}^i(t+1) = X^i(t) + sum_{j \in J_2} weight_j(speed) speedratio_j X_1^j(t+1) \tag{5.2.15}$$

$$\hat{Y}^i(t+1) = Y^i(t) + sum_{j \in J_2} weight_j(speed) speedratio_j Y_1^j(t+1) \tag{5.2.16}$$

Algorithm Four:

Assumption: The momentum for hurricane i stays the same from t to $t+1$. Momentum means speed and direction.

$$X^i(t+1) = X^i(t) + X_1^i(t) \tag{5.2.17}$$

$$Y^i(t+1) = Y^i(t) + Y_1^i(t) \tag{5.2.18}$$

Algorithm Five:

Assumption: The mean of recent momentum for hurricane i is used as the momentum at $t+1$.

$$X^i(t+1) = X^i(t) + avg(X_1^i(s)) \tag{5.2.19}$$

$$Y^i(t+1) = Y^i(t) + avg(Y_1^i(s)) \tag{5.2.20}$$

s can take from 1 to k, or can be most recent time points, such as $\frac{k}{2}$ to k.

Algorithm Six:

Assumption: the speed and acceleration has a linear relationship with recent time points bearing more weights, and the relationship stays the same from t to $t+1$ for hurricane i.

Compute weighted least square regression of $X_1^i(t)$ on $X_1^i(t-1) + X_2^i(t-1)$ without the intercept and $Y_1^i(t)$ on $Y_1^i(t-1) + Y_2^i(t-1)$ without the intercept for $t = 3, \ldots, k$, with weights being $e^{-(k-t+1)}$. Denote the two coefficients to be β_{1X}, β_{2X} and β_{1Y}, β_{2Y}.

Compute prediction of hurricane i at time point $t+1$ to be

$$\hat{X}^i(t+1) = X^i(t) + \beta_{1X} X_1^i(t-1) + \beta_{2X} X_2^i(t-1) \tag{5.2.21}$$

$$\hat{Y}^i(t+1) = Y^i(t) + \beta_{1Y} Y_1^i(t-1) + \beta_{2Y} Y_2^i(t-1) \tag{5.2.22}$$

Algorithm Seven:

Assumption: the speed and acceleration relationship stays the same from t to $t+1$ for all hurricanes j, and direction is used for similarity.

For each hurricane j

Compute weighted least square regression of $X_1^j(t)$ on $X_1^j(t-1) + X_2^j(t-1)$ without intercept and $Y_1^j(t)$ on $Y_1^j(t-1) + Y_2^j(t-1)$ without intercept for $t = 3, \ldots, k$, with weights

$e^{-(k-t+1)}$. Denote the coefficients to be β_{j1X},β_{j2X} and β_{j1Y},β_{j2Y}.

Compute prediction of hurricane i at time point $t+1$ to be

$$\hat{X}^i(t+1) = X^i(t) + sum_{j\in J_1} weight_j(direction)(\beta_{j1X}X_1^j(t) + \beta_{j2X}X_2^j(t)) \tag{5.2.23}$$

$$\hat{Y}^i(t+1) = Y^i(t) + sum_{j\in J_1} weight_j(direction)(\beta_{j1Y}Y_1^j(t) + \beta_{j2Y}Y_2^j(t)) \tag{5.2.24}$$

5.3 Combining Multiple Predictions

Once we have the initial predictions based on the seven individual algorithms, which are denoted as $\hat{X}^i(t)^{(n)}$ and $\hat{Y}^i(t)^{(n)}$, where $n = 1, 2, \ldots, 7$, we need to combine those individual results into our final prediction $\hat{Y}^i(t)^f$ and $\hat{Y}^i(t)^f$. In this section, we propose fifteen methods for computing the final prediction, and use euclidian distance to measure the prediction error: For prediction error at time t, $e(t)^f = \sqrt{(\hat{X}^i(t)^f - X^i(t))^2 + (\hat{Y}^i(t)^f - Y^i(t))^2}$.

Method One-Seven: Use Algorithm One to Seven respectively for sequential update. To be specific, for each data point $t = k+1, \ldots, k+m$, perform one step prediction from t to $t+1$, and obtain $\hat{Y}^i(t+1)^{(n)}$ where $n = 1, 2, \ldots, 7$ corresponding to Algorithm One to Seven.

Method Eight: Combine the prediction results One to Seven with equal weight $w(t)^{(n)} = 1/7$, and compute a weighted average for final prediction.

$$\hat{X}^i(t+1)^f = \sum_{n=1}^{7} \frac{1}{7}\hat{X}^i(t)^{(n)} \tag{5.3.25}$$

$$\hat{Y}^i(t+1)^f = \sum_{n=1}^{7} \frac{1}{7}\hat{Y}^i(t)^{(n)} \tag{5.3.26}$$

Method Nine: Use Bayes update. Denote the prediction error of each individual algorithm n at time t to be $e(t)^{(n)} = \sqrt{(\hat{X}(t)^{(n)} - X(t))^2 + (\hat{Y}(t)^{(n)} - Y(t))^2}$. We first initialize weights $w(k)^{(n)} = \frac{1}{7}$. At each time point t, where $t = k+1, \ldots, k+m$, update the weights based on the prediction error $e(t)^{(n)}$.

$$w(t+1)^{(n)} = w(t)^{(n)} \exp\{-500e(t)^{(n)}\} \tag{5.3.27}$$

Method Ten: Use Bayes update, but instead update the weight based on ${e(t)^{(n)}}^2$.

$$w(t+1)^{(n)} = w(t)^{(n)} \exp\{-500(e(t)^{(n)})^2\} \tag{5.3.28}$$

Method Eleven: Use Bayes update, but use 100 instead of 500 in Method Nine.

$$w(t+1)^{(n)} = w(t)^{(n)} \exp\{-100e(t)^{(n)}\} \tag{5.3.29}$$

Method Twelve: Use Bayes update, but use 100 instead of 500 in Method Ten.

$$w(t+1)^{(n)} = w(t)^{(n)} \exp\{-100(e(t)^{(n)})^2\} \tag{5.3.30}$$

Method Thirteen: Use Bayes update, but only assign weight of 1/2 to the two algorithms with the two smallest prediction errors in the previous time point. Specifically, assume that in time point t, $e(t)^{(j)}$ and $e(t)^{(k)}$ are the two smallest, then $w(t+1)^{(j)} = w(t+1)^{(k)} = 1/2$, and $w(t+1)^{(n)} = 0$ for $n \neq j, k$.

Method Fourteen: Use Bayes update, but only assign weight of 1/2 to the two algorithms with the two smallest cumulative prediction errors in the previous two time points. Specifically, assume that $e(t-1)^{(j)} + e(t)^{(j)}$ and $e(t-1)^{(k)} + e(t)^{(k)}$ are the two smallest, then $w(t+1)^{(j)} = w(t+1)^{(k)} = 1/2$, and $w(t+1)^{(n)} = 0$ for $n \neq j, k$.

Method Fifteen: Use Bayes updates, set initial weight $w(t)^{(n)} = 1/7$, sort $e(t)^{(n)}$ from smallest to largest and assign rank $R^{(n)}$. Update $w(t)^{(n)}$ with the following:

$$w(t+1)^{(n)} = w(t)^{(n)} + \frac{1}{21}(4 - R^{(n)}) \tag{5.3.31}$$

This may generate negative weights.

5.4 Multiple Step Ahead Direct Prediction

In this section, we extend the m step ahead sequential prediction to m step ahead direct prediction. To evaluate the performance of Method 1-15, we need to split the observations into three components: training data, validation data and testing data.

For the training data, we select data points from $t = 1$ to k. It is mainly used in two ways: firstly, we may find similar hurricanes based on previous direction or speed; therefore we can obtain a candidate set for collaborative filtering approach. Secondly, we may leverage past hurricane speed information to project onto the target hurricane.

For the validation data, we select data points from $t = k + 1$ to $k + r$. This is used for weight updating, which is used in Method 9-15. In each method, we leverage the sequential prediction procedures in Section 5.3 by computing the prediction error for every t and updating weights for each algorithm. At time $t = k + r$, we will have updated weights for every algorithm. Note that in Algorithm 1-8, the weights are unchanged.

For the test data, we select data points from $t = k + r + 1$ to $k + r + m$. This is used for evaluating the performance of the fifteen proposed methods. We report the mean squared prediction error for each method.

5.5 Result and Evaluation

In this section, we show two example hurricanes: 1952 2nd hurricane and 2008 4th hurricane. In the first part, we examine the effectiveness of different algorithms under the sequential prediction context, which is shown in Scenario 1-5. In the second part, we compare the prediction accuracy between m step ahead sequential prediction and m step ahead direct prediction, which is shown in Scenario 6-9. We choose $m = 4$ for illustration.

Scenario One: 2nd hurricane in 1952: We observe the initial 8 data points, and want to predict the 9th point until the 20th point sequentially. Please refer to Table 5.1.

Scenario Two: 2nd hurricane in 1952: We observe the initial 8 data points, and want to predict the 9th point until the 40th point sequentially. Please refer to Table 5.2.

Scenario Three: 2nd hurricane in 1952: We observe the initial 12 data points, and want to predict the 13th point until the 30th point sequentially. Please refer to Table 5.3.

Scenario Four: 4th hurricane in 2008: We observe the initial 8 data points, and want to predict the 9th point until the 20th point sequentially. Please refer to Table 5.4.

Scenario Five: 4th hurricane in 2008: We observe the initial 12 data points, and want to predict the 13th point until the 25th point sequentially. Please refer to Table 5.5.

Scenario Six: 2nd hurricane in 1952: We observe the initial 12 data points(8 points for training, 4 points for validation), and want to predict the 13th point until the 16th point. Please refer to Table 5.6 and Figure 5.1.

Scenario Seven: 2nd hurricane in 1952: We observe the initial 16 data points(12 points for training, 4 points for validation), and want to predict the 17th point until the 20th point. Please refer to Table 5.7 and Figure 5.2.

Scenario Eight: 4th hurricane in 2008: We observe the initial 12 data points(8 points for training, 4 points for validation), and want to predict the 13th point until the 16th point. Please refer to Table 5.8 and Figure 5.3.

Scenario Nine: 4th hurricane in 2008: We observe the initial 16 data points(12 points for training, 4 points for validation), and want to predict the 17th point until the 20th point. Please refer to Table 5.9 and Figure 5.4.

From the table above, we have the following observations: Firstly, Algorithm Four is consistently among the best in sequential prediction. This is reasonable because in one step ahead prediction, it is natural to assume that momentum persists, i.e., speed and direction remain the same in the next 6 hours. Secondly, Algorithm Eleven is consistently outstanding as well. Dynamic weighting scheme has been promising as it quickly adapts to the new data and smartly learns how to best aggregate different model results. Thirdly, algorithms with collaborative filtering such as Algorithm One have dramatic prediction improvement in the 2008 hurricane compared to the 1952 hurricane. This is not surprising because with collaborative filtering, we need to leverage past hurricane's trajectory information. As we take more trajectory information from past hurricanes into consideration, this method becomes more effective.

5.6 Discussion and Future Work

Although our methods have shown promising results, there is some future work which requires attention.

One area of improvement is to incorporate parametric curve fitting. Most Atlantic hurricanes start in the center of the Atlantic ocean heading westbound. At a certain point in time due to land or climate effect, they curve northeast bound, which exhibits a rotated U shaped curve. This inspires us to fit a quadratic curve $f(x, y) = \epsilon$. We have tried five models for quadratic curve fitting: the degenerated linear model, standard parabola along x and y axis, general parabola and general quadratic models. Each model fits a curve, and by the goodness of fit criteria, we may assign the weight of each curve. With the additional assumption that ϵ follows Gaussian distribution with mean zero, we may derive BIC value for model selection purposes.

Another direction is to perform regression with climate covariates. We have collected North Atlantic Oscillation (NAO), Southern Oscillation Index (SOI), Sea Surface Temperature (SST), Carbon Dioxide (CO2) and El Nino/La Nina data as covariates, and we have tried a variety of hurricane characteristics. One promising result is that the second derivatives for latitude and longitude seem to follow bivariate normal distribution from various plots we have drawn, and we have modeled the correlation of the second derivatives of X and Y as a linear function of climate covariates. The final regression model has an R square of 0.32. We feel the area of climate covariate modeling may have more promising results as we dive in.

The last area of research we would like to note is the lagging logic for collaborative filtering. We have showed one approach of implementing collaborative filtering by comparing the first k points of past hurricanes and the target hurricane. We may extend the concept of similar hurricane by including the lagging logic as some hurricanes might behave similarly as the target hurricane, but starting from the middle of its life. The candidate hurricane set could potentially include more quality hurricanes used for prediction purposes.

Method	Mean Err $t = 9 - 20$	Mean Err $t = 15 - 20$	Sd Dev $t = 9 - 20$	Sd Dev $t = 15 - 20$
1	1.8073959	1.8739839	0.3128219	0.4119217
2	1.6480342	1.7159472	0.2123641	0.2579312
3	1.3315665	1.2474672	0.2945322	0.3990878
4	**0.2706130**	**0.3454164**	0.1412222	0.1575570
5	0.6611576	0.8646424	0.2753069	0.2133239
6	0.4677332	0.6176777	0.3338800	0.4198622
7	2.4376227	2.7626841	0.7541810	0.9357994
8	0.8269695	0.8182070	0.1639804	0.2210493
9	**0.3322354**	**0.3515392**	0.1929289	0.1563864
10	0.7430370	0.6967022	0.1586472	0.2012450
11	**0.4135457**	**0.4214280**	0.2008046	0.1973591
12	0.8109716	0.7957325	0.1593774	0.2135730
13	0.4659567	0.5807229	0.2561372	0.2678512
14	0.5319953	0.5848100	0.2253086	0.2018914
15	2.5015669	4.0757204	2.3605839	2.4516261

Table 5.1: Prediction Error for 1952 2nd Hurricane with 8 initial points and sequential prediction until the 20th point.

Method	Mean Err $t = 9 - 40$	Mean Err $t = 25 - 40$	Sd Dev $t = 9 - 40$	Sd Dev $t = 25 - 40$
1	1.4511240	1.1755662	0.5619302	0.6153152
2	1.4703772	1.1953514	0.6721736	0.8359838
3	1.4143571	1.3640513	0.5328540	0.6289583
4	**0.2848747**	**0.3006806**	0.1658225	0.2003683
5	1.0524187	1.2880440	0.3742568	0.1893878
6	0.4436982	0.4671011	0.3886398	0.4669397
7	1.8059835	1.3317076	0.9705792	0.9061030
8	0.6426769	0.4964579	0.2836381	0.3019457
9	**0.3079948**	**0.3006806**	0.1840227	0.2003683
10	0.5039008	0.3597985	0.2659458	0.2260799
11	**0.3474711**	**0.3017559**	0.1934839	0.2007951
12	0.5890129	0.4216438	0.2951691	0.2947838
13	0.4950577	0.5415632	0.4110281	0.5245919
14	0.5696212	0.6213473	0.4153983	0.5472561
15	6.3549820	9.3612426	4.4706315	3.8959221

Table 5.2: Prediction Error for 1952 2nd Hurricane with 8 initial points and sequential prediction until the 40th point.

Method	Mean Err $t = 13 - 30$	Mean Err $t = 22 - 30$	Sd Dev $t = 13 - 30$	Sd Dev $t = 22 - 30$
1	1.6624758	1.4880355	0.4137979	0.4313028
2	1.7428801	1.7769426	0.3467666	0.4471603
3	1.5602813	1.7714985	0.4698372	0.4811223
4	**0.3433010**	**0.3705711**	0.1558775	0.1797166
5	1.1150558	1.3739376	0.3369047	0.1744682
6	0.5493102	0.5651113	0.4817371	0.6035833
7	2.0927237	1.7278157	0.8874645	0.7342394
8	0.7489626	0.7017779	0.2087491	0.2288190
9	**0.3908820**	**0.3705785**	0.2072778	0.1797115
10	0.5615858	0.3986808	0.2642044	0.2222984
11	**0.4257411**	**0.3870857**	0.1858850	0.1743434
12	0.6854019	0.5905768	0.2408749	0.2608838
13	0.5147449	0.4205346	0.2784364	0.2547802
14	0.5882483	0.5299529	0.2506130	0.2985262
15	3.7379878	6.0777058	3.0441479	2.5042988

Table 5.3: Prediction Error for 1952 2nd Hurricane with 12 initial points and sequential prediction until the 30th.

Method	Mean Err $t = 9 - 20$	Mean Err $t = 15 - 20$	Sd Dev $t = 9 - 20$	Sd Dev $t = 15 - 20$
1	0.4525609	0.5296689	0.2643484	0.3586558
2	0.4192896	0.4643861	0.2071213	0.2690761
3	0.7282678	0.9156530	0.3417735	0.3896244
4	**0.2875578**	**0.3193387**	0.1726671	0.1440332
5	0.5448656	0.4488279	0.1716571	0.1894056
6	0.4338514	0.5619376	0.2544866	0.2417623
7	0.4843067	0.4354321	0.1974556	0.2556198
8	0.3298292	0.3936615	0.1836887	0.2446580
9	**0.3131479**	**0.3090476**	0.1310436	0.1720148
10	0.3217051	0.3755547	0.1773621	0.2400345
11	**0.2959056**	**0.3085983**	0.1633193	0.2314414
12	0.3280264	0.3896892	0.1821320	0.2434125
13	0.3600642	0.4189966	0.1799547	0.2179782
14	0.3195816	0.3627734	0.1679452	0.2214300
15	0.5227889	0.7195175	0.4610437	0.6062755

Table 5.4: Prediction Error for 2008 4th Hurricane with 8 initial points and sequential prediction until the 20th point.

Method	Mean Err $t = 12 - 25$	Mean Err $t = 12 - 25$	Sd Err $t = 19 - 25$	Sd Err $t = 19 - 25$
1	0.4380975	0.5231358	0.2859269	0.2919156
2	0.3679111	0.3469852	0.2104851	0.1639306
3	0.7722712	0.9029711	0.3899604	0.4428410
4	**0.2849047**	0.2733404	0.1508025	0.1656224
5	0.4836109	0.5342791	0.1426544	0.1173311
6	0.4063773	0.4568269	0.2516215	0.2971570
7	0.4230711	0.4704076	0.1961908	0.1904797
8	0.3320027	0.3585199	0.1994198	0.1751078
9	0.2918306	**0.2728033**	0.1840498	0.1621946
10	0.3219593	0.3405278	0.1895065	0.1563697
11	**0.2838750**	**0.2719790**	0.1725412	0.1266155
12	0.3296582	0.3543044	0.1970230	0.1706731
13	0.3229364	0.3620306	0.1773588	0.1868320
14	**0.2915537**	**0.2682794**	0.1631240	0.1127745
15	0.5853016	0.8448210	0.4271994	0.4040738

Table 5.5: Prediction Error for 2008 4th Hurricane with 12 initial points and sequential prediction until the 25th point.

Method No.	Mean Err Four Step $t = 13 - 16$	Mean Error Sequential $t = 13 - 16$
1	4.1777467	1.7758991
2	3.7682447	1.6564291
3	4.0105785	1.3222466
4	1.1029331	**0.2246270**
5	2.0006046	0.6736875
6	0.6584243	0.2707242
7	5.4228702	2.0442073
8	2.2597505	0.7168782
9	1.1280290	0.3687886
10	1.9493101	0.7238374
11	**0.6542515**	0.4699195
12	2.1942733	0.7191414
13	0.8744275	0.4590143
14	1.5232310	0.5737462

Table 5.6: Comparison of the prediction Error for 1952 2nd Hurricane with 12 initial points and prediction from 13th until the 16th point. The second column is the mean prediction error by four step ahead prediction, and the third column is the mean prediction error by sequential prediction. Method 15 is not shown due to the non convexity of the weights.

Method	Mean Err Four Step $t = 17 - 20$	Mean Error Sequential $t = 17 - 20$
1	4.6758065	1.9083910
2	3.6520192	1.7195374
3	2.9911712	1.2596737
4	0.9489624	**0.4294246**
5	2.6405440	0.9423072
6	0.4921466	0.8115585
7	6.1028942	3.1366829
8	1.7719443	0.9084986
9	0.9478171	0.6657341
10	1.7924217	0.8267935
11	**0.4577258**	0.6084452
12	1.7846199	0.8891580
13	0.6471807	0.8073784
14	1.7208895	0.8902328

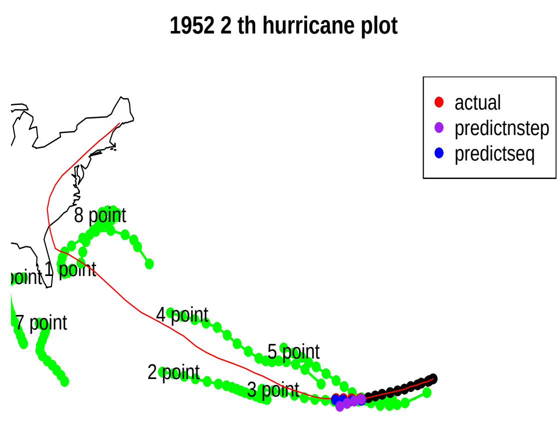
Table 5.7: Comparison of the prediction Error for 1952 2nd Hurricane with 16 initial points and prediction from 17th until the 20th point.

Method	Mean Err Four Step $t = 13 - 16$	Mean Error Sequential $t = 13 - 16$
1	0.4548712	0.2068796
2	0.5957023	0.2603097
3	1.1087301	0.4934851
4	0.2603326	0.2308824
5	1.4554027	0.4325193
6	0.3734506	0.2546254
7	0.6401927	0.2764625
8	0.2233624	0.1814977
9	**0.2195195**	0.2012980
10	0.2231173	0.1811979
11	0.2328813	**0.1800382**
12	0.2230923	0.1814364
13	0.4639136	0.1910183
14	0.2989723	0.2086916

Table 5.8: Comparison of the prediction Error for 2008 4th Hurricane with 12 initial points and prediction from 13th until the 16th point.

Method	Mean Err Four Step $t = 17 - 20$	Mean Error Sequential $t = 17 - 20$
1	1.5231079	0.7208676
2	1.8889045	0.5927748
3	2.4657394	1.0952139
4	1.4857458	**0.3536876**
5	0.9476788	0.5190396
6	**0.9781338**	0.6780547
7	1.2467078	0.5797154
8	1.4603374	0.5135106
9	1.4284063	0.4048431
10	1.4547345	0.4989895
11	1.4130335	0.4317280
12	1.4591776	0.5104222
13	1.5011300	0.5276454
14	1.3650427	0.4407171

Table 5.9: Comparison of the prediction Error for 2008 4th Hurricane with 16 initial points and prediction from 17th until the 20th point.



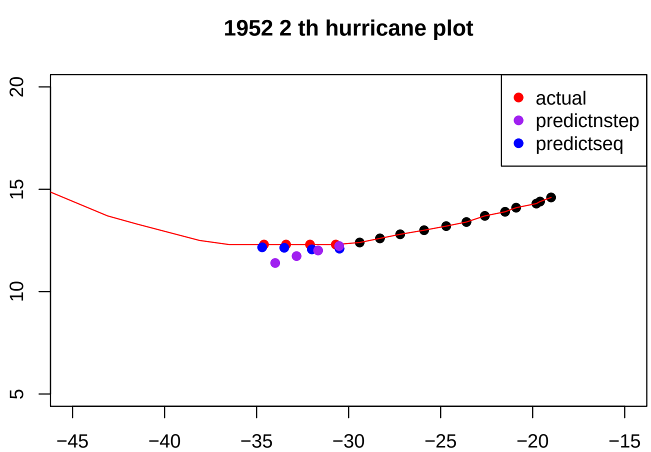
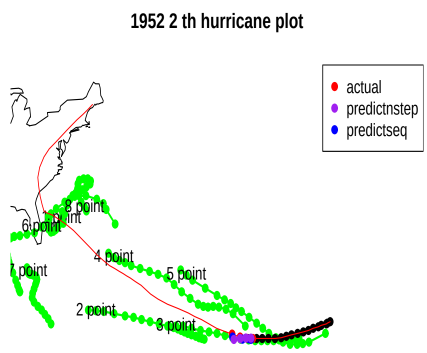


Figure 5.1: 1952 2nd hurricane prediction with 12 initial points and predict the 13th till 16th point.



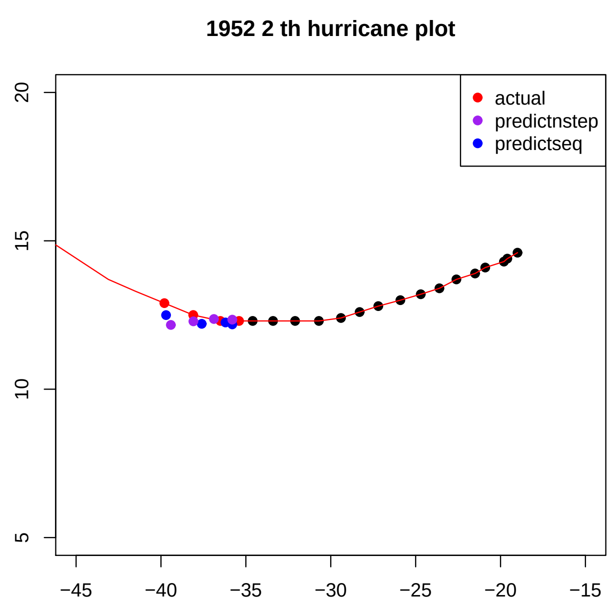
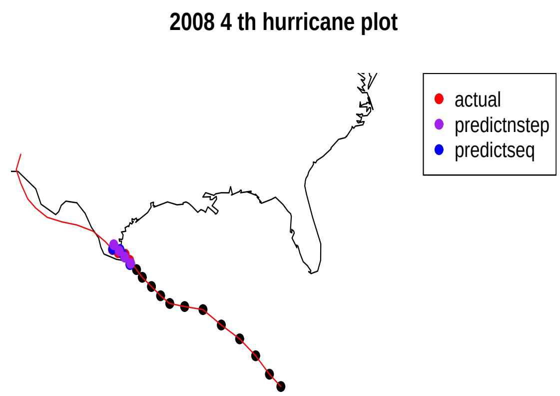


Figure 5.2: 1952 2nd hurricane prediction with 16 initial points and predict the 17th till 20th point.



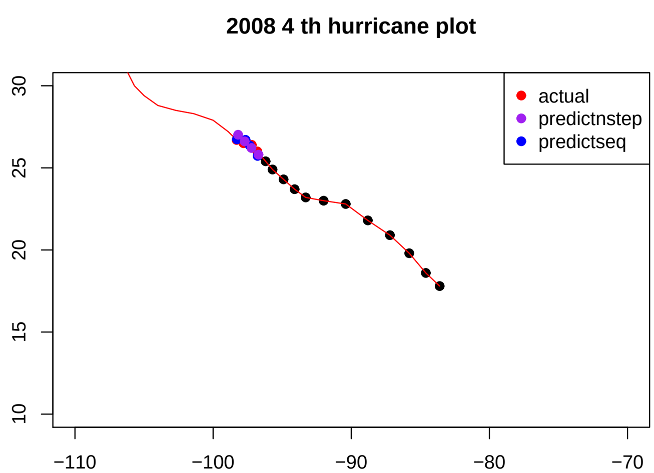
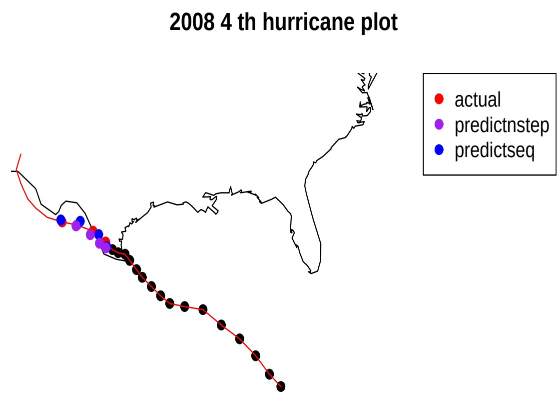


Figure 5.3: 2008 4th hurricane prediction with 12 initial points and predict the 13th till 16th point.



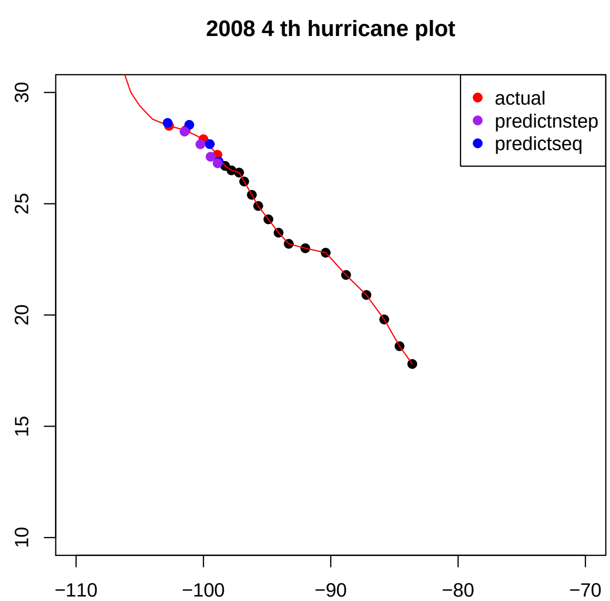


Figure 5.4: 2008 4th hurricane prediction with 16 initial points and predict the 17th till 20th point.

Bibliography

[1] Agresti, A. and Coull, B.A. (1998) Approximate Is Better Than 'Exact' for Interval Estimation of Binomial Proportions. *The American Statistician*, **52**, 119-126.

[2] Akaike, H. (1974) A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, **19(6)**, 716-723.

[3] Albers, W. and Kallenberg, W.C.M. (2009) CUMIN Charts. *Metrika*, **70**, 111-130, 2009.

[4] Alvarado, E., Sandberg, V.D. and Pickford, G.S. (1998) Modeling Large Forest Fires as Extreme Events. *Northwest Science*, **72**, 66-75, 1998.

[5] Alwan,L.C. (2000) Designing an Effective Exponential CUSUM Chart without the Use of Nomographs. *Communications in Statistics-Theory and Methods*, **29(12)**, 2879-2893, 2000.

[6] Atienza, O.O., Tang, L.C., Ang, B.W. (2000) A Uniform Most Powerful Cumulative Sum Scheme Based on Symmetry. *Journal of the Royal Statstical Society. Series D*, **49(2)**, 209-217, 2000.

[7] Barratt, B., Atkinson, R., Anderson, H.R., Beevers, S., Kelly, F., Mudway, I. and Wilkinson, P. (2007) Investigation into the Use of the CUSUM Technique in Identifying Changes in Mean Air Pollution Levels Following Introduction of a Traffic Management Scheme. *Atmospheric Environment*, **41(8)**, 1784-1791, 2007.

[8] Blondell, D., Hoang, P., Powell, G.J. and Shi, J. (2002) Detection of Financial Time Series Turning Points: A New CUSUM Approach Applied to IPO Cycles. *Review of Quantitative Finance and Accounting*, **18(3)**, 293-315, 2002.

[9] Bolton, R.J. and Hand, D.J. (2002) Statistical fraud detection: A Review. *Statistical Science*, **17**, 235-255, 2002.

[10] Bon, K.S. (2007) Probability Model Selection Using Information-Theoretic Optimization Criterion. *Journal of Statistical Computation and Simulation*, **69(3)**, 2001.

[11] Borror, M.C., Montgomery, C.D. and Runger, C.G. (1999) Robustness of the EWMA Control Chart to Non-normality. *Journal of Quality Technology*, **31(3)**, 309-316, 1999.

[12] Brown, R.L., Durbin, J. and Evans, J.M. (1975) Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistical Society. Series B (Methodological)*, **37(2)**, 149-192, 1975.

[13] Castillo, E. (1988) *Extreme Value Theory in Engineering*. Academic Press, Inc. New York.

[14] Lu,Y. and Chatterjee,S. (2014) Instability and change detection in exponential families and generalized linear models, with a study of Atlantic tropical storms. *Nonlin. Processes Geophys.*,**21**, 1133-1143, 2014.

[15] Lu,Y. and Chatterjee,S. (2015) Instability and change detection in extremes, with a study of Atlantic tropical storms. *Preprint*

[16] Chatterjee, S. and Qiu, P. (2009) Distribution-Free Cumulative Sum Control Charts Using Bootstrap-Based Control Limits. *The Annals of Applied Statistics*, **3(1)**, 349-369, 2009.

[17] Chen, M.Y. (2002) Tests for Parameter Constancy in Regression Models. 2002.

[18] Chen, J. and Gupta, A.K. (1997) Testing and Locating Variance Changepoints with Application to Stock Prices. *Journal of the American Statistical Association*, **92(438)**, 739-747, 1997.

[19] Chen, J., Gupta, A. K., Pan, J. (2006) Information Criterion and Change Point Problem for Regular Models. *The Indian Journal of Statistics*, **68**,252-282, 2006.

[20] Chihwa, K. and Ross, S.L. (1995). A Cusum Test in the Linear Regression Model with Serially Correlated Disturbances. *Econometric Reviews*, **14(3)**, 331-346, 1995.

[21] Chen, W., Chen, D., Sun, G. and Zhang, Y. (2007) Defending Against Jamming Attacks in Wireless Local Area Networks. *Lecture Notes in Computer Science*, **4610**, 519-528, 2007.

[22] Coles, S. (2001) An Introduction to Statistical Modeling of Extreme Values. Springer.

[23] Crosier,R.B. (1988) Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics*, **30(3)**, 291-303, 1988.

[24] Dacorogna, M.M., Pictet, O.V., Müller, U.A. and de Vries, C.G. (2001) Extremal Forex Returns in Extremely Large Data Sets. *Extremes*, **4**, 105-127, 2001.

[25] Danielsson, J. and de Vries, C.G. (1997) Tail Index and Quantile Estimation with Very High Frequency Data. *Journal of Empirical Finance*, **4**, 241-257, 1997.

[26] Deng, A. and Perron, P. (2008) The Limit Distribution of the CUSUM of Squares Test under General Mixing Conditions. *Econometric Theory*, 2008.

[27] Devijver, P. A. and Kittler, J. (1982) Pattern Recognition: A Statistical Approach. *Prentice-Hall, London*.

[28] Elsner, J.B. (2003) Tracking Hurricanes. *Bulletin of the American Meteorological Society*, **84**, 353-356, 2003.

[29] Elsner, J.B. (2006) Evidence in Support of the Climate Change-Atlantic Hurriane Hypothesis. *Geophysical Research letters*, **33**, L16705, 2006.

[30] Elsner, J.B. (2007) Granger Causality and Atlantic Hurricanes. *TellusA*, **59(4)**, 476-485, 2007.

[31] Elsner, J.B., and Bossak, B.H. (2001) Bayesian Analysis of U.S. Hurricane Climate. *Journal of Climate*, **14**, 4341-4350, 2001.

[32] Elsner, J.B., Bossak, B.H. and Niu, X. (2001) Secular Changes to the ENSO-U.S. Hurricane Relationship. *Geophysical Research Letters*, **28**, 4123-4126, 2001.

[33] Elsner, J. B., and Bossak, B.H. (2004) Hurricane Landfall Probability and Climate. *Hurricanes and Typhoons: Past, Present, and Future*, Columbia University Press.

[34] Elsner, J.B. Hodges, R.E. and Jagger, T.H. (2011) Climate Dynamics, August 2011.

[35] Elsner, J.B. and Jagger, T.H. (2004) A Hierarchical Bayesian Approach to Seasonal Hurricane Modeling. *Journal of Climate*, **17**, 2813-2827, 2004.

[36] Elsner, J.B. and Jagger, T.H. (2006) Prediction Models for Annual U.S. Hurricane Counts. *Journal of Climate*, **19**, 2935-2952, 2006.

[37] Elsner, J.B. and Jagger, T.H. (2006) Comparison of Hindcasts Anticipating the 2004 Florida Hurricane Season. *Weather and Forecasting*, **21**, 182-192, 2006.

[38] Elsner, J.B. and Jagger, T.H. (2008) United States and Caribbean Tropical Cyclone Activity Related to the Solar Cycle. *Geophysical Research Letters*, **35**, L18705, 2008.

[39] Elsner, J.B. and Jagger, T.H. (2008) Improving Multiseason Forecasts of North Atlantic Hurricane Activity. *Journal of Climate*, **21**, 1209-1219, 2008.

[40] Elsner, J.B., Jagger, T.H. and Fogarty, E.A. (2009) Visibility Network of United States Hurricanes. *Geophysical Research Letters*, **36**, L16702, 2009.

[41] Elsner, J.B., Jagger, T.H. and Niu, X. (2000) Changes in the Rates of North Atlantic Major Hurricane Activity During the 20th Century. *Geophysical Research Letters*, **27**, 1743-1746, 2000.

[42] Elsner, J.B., Jagger, T.H. and Hodges, R.E. (2010) Daily Tropical Cyclone Intensity Response to Solar Ultraviolet Radiation, *Geophysical Research Letters.* **37**, L09701, 2010.

[43] Elsner, J.B., Jagger, T.H. and Tsonis, A.A. (2006) Estimated Return Periods for Hurricane Katrina. *Geophysical Research Letters*, **33**, 08704, 2006.

[44] Elsner, J.B. and Kavlakov, S.P. (2001) Hurricane Intensity Changes Associated with Geomagnetic Variation. *Atmospheric Science Letters*, **2**, 86-93, 2001.

[45] Elsner, J.B. and Kocher, B. (2000) Global Tropical Cyclone Activity: A Link to the North Atlantic Oscillation. *Geophysical Research Letters*, **27**, 129-132, 2000.

[46] Elsner, J.B., Kossin, J.P. and Jagger, T.H. (2008) The Increasing Intensity of the Strongest Tropical Cyclones. *Nature*, **455**, 92-95, 2008.

[47] Elsner, J.B., Liu, K. and Kocher, B. (2000) Spatial Variations in Major U.S. Hurricane Activity: Statistics and a Physical Mechanism. *Journal of Climate*, **13**, 2293-2305, 2000.

[48] Elsner, J.B., Murnane, R.J. and Jagger, T.H. (2006) Forecasting US Hurricanes 6 Months in Advance. *Geophysical Research Letters*, **33**, L10704, 2006.

[49] Elsner, J.B., Niu, X. and Jagger, T.H. (2004) Detecting Shifts in Hurricane Rates Using a Markov Chain Monte Carlo Approach. *Journal of Climate*, **17**, 2652-2666, 2004.

[50] Elsner, J.B., Niu, X. and Tsonis, A.A. (1998) Multi-year Prediction Model of North Atlantic Hurricane Activity. *Meteorology and Atmospheric Physics*, **68**, 43-51, 1998.

[51] Elsner, J.B., Tsonis, A.A. and Jagger, T.H. (2006) High Frequency Variability in Hurricane Power Dissipation and Its Relationship to Global Temperature. *Bulletin of the American Meteorological Society*, **87**, 763-768, 2006.

[52] Geisser, S. (1993) Predictive Inference. *New York: Chapman and Hall*.

[53] Gilli, M. and Këllezi, E. (2006) An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics*, **27(1)**, 1-23, 2006.

[54] Gombay, E. and Horvath, L. (1990) Asymptotic Distributions of Maximum Likelihood Tests for Change in the Mean. *Biometrika*, **77(2)** 411-414.

[55] Hauksson, H.A., Dacorogna, M., Domenig, T., Müller, U. and Samorodnitsky, G. (2000) Multivariate Extremes, Aggregation and Risk Estimation. *Manuscript. School of Operations Research and Industrial Engineering, Cornell University*.

[56] Hawkins, D.M. (1977) Testing a Sequence of Observations for a Shift in Location. *Journal of American Statistical Association*, **72**, 180-186.

[57] Hawkins, D.M. (1992) Evaluation of Average Run Lengths of Cumulative Sum Charts for an Arbitrary Data Distributions. *Communications in Statistics-Simulation and Computation*, **21(4)**, 1001-1020, 1992.

[58] Hawkins, D.M. and Olwell, H.D. (1998) Cumulative Sum Charts and Charting for Quality Improvement. *Springer-Verlag*, New York, 1998.

[59] Hawkins, D.M. and Olwell, D.H. (1997) Inverse Gaussian Cumulative Sum Control Charts for Location and Shape. *Journal of the Royal Statistical Society. Series D*, **46(3)**, 323-335, 1997.

[60] Hawkins, D.M. and Zamba, K.D. (2005) A Change-point Model for a Shift in Variance. *Journal of Quality Technolology*, **37(1)**, 21-31, 2005.

[61] Healy,J.D. (1987) A Note on Multivariate CUSUM Procedures. *Technometrics*, **29(4)**, 409-412, 1987.

[62] Horvath L. et al (2008) Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen. *IMS collections*, **1**, 293-304, 2008.

[63] Hess, J.C., Elsner, J.B. and Laseur, N.E. (1995) Improving Seasonal Hurricane Predictions for the Atlantic Basin. *Weather and Forecasting*, **10**, 425-432, 1995.

[64] Hinkley, D.V. and Hinkley, E.A. (1970) Inference about the Change-Point in a Sequence of Binomial Variables. *Biometrika*, **57(3)**, 477-488.

[65] Hodges, R.E. and Elsner, J.B. (2010) Evidence Linking Solar Variability with US Hurricanes. *International Journal of Climatology*, doi:10.1002/joc.2196, 2010.

[66] Horvath, L. and Serbinowska, M. (1995) Testing for Changes in Multinomial Observations: The Lindisfarne Scribes Problem. *Scandinavian Journal of Statistics*, **22(3)**, 371-384.

[67] Hsu, D.A. (1979) Detecting Shifts of Parameter in Gamma Sequences with Applications to Stock price and Air Traffic Flow Analysis. *Journal of American Statistical Association*, **74**, 31-46.

[68] Jagger, T.H. and Elsner, J.B. (2008) Modeling Tropical Cyclone Intensity with Quantile Regression. *International Journal of Climatology*, doi:10.1002/joc.1804, 2008.

[69] Jagger, T.H. and Elsner, J.B. (2010) A Consensus Model for Seasonal Hurricane Prediction. *Journal of Climate*, **23**, 6090-6099, 2010.

[70] Jagger, T.H., Elsner, J.B. and Niu, X. (2001) A Dynamic Probability Model of Hurricane Winds in Coastal Counties of the United States, *Journal of Applied Meteorology*, **40**, 853-863, 2001.

[71] Jagger, T.H., Elsner, J.B. and Sounders, M.A. (2008) Forecasting U.S. Insured Hurricane Losses. *Climate Extremes and Society*, Cambridge University Press, 189-208, 2008.

[72] Jagger, T.H., Niu, X. and Elsner, J.B. (2002) A Space-Time Model for Seasonal Hurricane Prediction. *International Journal of Climatology*, **22**, 451-465, 2002.

[73] Jandhyala, V.K. and MacNeill, I.B. (1991) Tests for Parameter Changes At Unknown Times in Linear Regression Models. *Journal of Statistical Planning and Inference*, **27(3)**, 291-316, 1991.

[74] Jandhyala,V.K. and Al-Saleh,A.J. (1999) Parameter Changes at Unknown Times in Non-Linear Regression. *Environmetrics*, **10(6)**, 711-724, 1999.

[75] Johnson, N.L. (1966) Cumulative Sum Control Charts and the Weibull Distribution. *Technometrics*, **8(3)**, 481-491, 1966.

[76] Kahya, E. and Theodossiou, P. (1999) Predicting Corporate Financial Distress: A Time-Series CUSUM Methodology. *Review of Quantitative Finance and Accounting*, **13**, 323-345, 1999.

[77] Khan, R.A. (1979) A Sequence Detection Procedure and the Related Cusum Procedure. *The Indian Journal of Statistics, Series B*, **40(3/4)**, 146-162, 1979.

[78] Kim, M. and Lee, S. (2009) Test for Tail Index Change in Stationary Time Series with Pareto-type Marginal Distribution. *Bernoulli*, **15(2)**, 325-356, 2009.

[79] Koedijk, K.G., Schafgans, M.M.A. and de Vries, C.G. (1990) The Tail Index of Exchange Rate Returns. *Journal of International Economics*, **29**, 93-108, 1990.

[80] Kohavi, Ron (1995) A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, **2(12)**, 1137-1143.

[81] Krawczak,M., Ball,E., Fenton,I., Stenson,P., Abeysinghe,S., Thomas,N. and Cooper,D.N. (1999) Human Gene Mutation Database: A Biomedical Information and Research Resource.*Human Mutation*,**15**, 45-51, 1999.

[82] Krawczak, M., Ball, E., Fenton, I., Stenson, P., Abeysinghe, S., Thomas, N. and Cooper, D. N. (1999). Human Gene Mutation Database: A biomedical Information and Research Resource.*Human Mutation*,**15**, 45C51, 1999.

[83] Leahy, P.G. and Kiely, G. (2011) Short Duration Rainfall Extremes in Ireland: Influence of Climatic Variability. *Water Resour Manage*, **25**, 987-1003, 2011.

[84] Lee, C.B.(1997) Estimating the Number of Change Points in Exponential Families Distributions. *Scandinavian Journal of Statistics*, **24**, 201-210.

[85] Lee, S., Ha, J., Na, O. and Na, S. (2003) The Cusum Test for Parameter Change in Time Series Models. *Scandinavian Journal of Statistics*, **30(4)**, 781-796, 2003.

[86] Lee, S., Tokutsu, Y. and Maekawa, K. (2004) The CUSUM Test for Parameter Change in Regression Models with ARCH Errors. *Journal of the Japanese Statistical Society*, **34(2)**, 173-188, 2004.

[87] Lehmiller, G.S., Kimberlain, T.B. and Elsner, J.B. (1997) Seasonal Prediction Models for North Atlantic Basin Hurricane Location, *Monthly Weather Review*, **125**, 1780-1791, 1997.

[88] Lim, T.O., Soraya, A., Ding, L.M. and Morad, Z. (2002) Assessing doctor's Competence: Application of CUSUM Technique in Monitoring Doctors' Performance. *International Journal for Quality in Health Care*, **14(3)**, 251-258, 2002.

[89] Liu, W.H. (2009) Detecting Structural Breaks in Tail Behavior-From the Perspective of the Generalized Pareto Distribution fitting. 2009.

[90] Liu, J.Y., Xie, M. and Goh, T,N. (2006) CUSUM Chart with Transformed Exponential Data. *Communications in Statistics-Theory and Methods*, **35(10)**, 1829-1843, 2006.

[91] Lorden,G. (1971) Procedures for Reacting to a Change in Distribution. *The Annals of Statistics*, **42(6)**, 1897-1908, 1971.

[92] Lucas, J.M. (1982) Combined Shewhart-CUSUM Quality Control Schemes. *Journal of Quality Technology*, **14**, 51-59, 1982.

[93] Lucas, M.J. (1985) Counted Data CUSUM's. *Technometrics*, **27(2)**, 129-144, 1985.

[94] Lucas,J.M.(1982) Combined Shewhart-CUSUM Quality Control Schemes. *Journal of Quality Technology*, **14**, 51-59, 1982.

[95] Lucas, M.J. and Crosier, R.B. (1982) Robust CUSUM: a Robustness Study for CUSUM Quality Control Schemes. *Communications in Statistics-Theory and Methods*, **11(23)**, 2669-2687, 1982.

[96] Lucas, J.M. and Saccucci, M.S. (1990) Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics*, **32(1)**, 1-12, 1990.

[97] Luger,R. (2001) A Modified CUSUM Test for Orthogonal Structural Changes. *Economics Letters*, **73(3)**, 301-306, 2001.

[98] MacEachern, S.N., Rao, Y. and Wu, C. (2007) A Robust-Likelihood Cumulative Sum Chart. *Journal of the American Statistical Association*, **102**, 1440-1447, 2007.

[99] Mallows, C.L. (1973) Some Comments on CP. *Technometrics*, **15(4)**, 661-675.

[100] Malmstadt, J., Scheitlin, K. and Elsner, J.(2007) Florida Hurricanes and Damage Costs. *Southeastern Geographer*, **49(2)**, 108-131, 2009.

[101] Malmstadt, J.C., Elsner, J.B. and Jagger, T.H. (2010) Risk of Strong Hurricane Winds to Florida Cities. *Journal of Applied Meteorology and Climatology*, **49**, 2121-2132, 2010.

[102] Manly, F.J.B and Mackenzie, I.D. (2003) CUSUM Environmental Monitoring in Time and Space. *Environmental And Ecological Statistics*, **10(2)**, 231-24, 2003.

[103] McGilchrist, C.A. and Matawie, K.M. (1998) Recursive Residuals in Generalised Linear Models. *Journal of Statistical Planning and Inference*, **70**, 335-344, 1998.

[104] McNees, P., Dow, K.H. and Loerzel, V.W. (2005) Application of the CUSUM Technique to Evaluate Changes in Recruitment Strategies. *Nursing Research*, **54(6)**, 399-405, 2005.

[105] Meulbroek, L.K. (1992) An Empirical Analysis of Illegal Insider Trading. *Journal of Finance*, **47**, 1661-1699, 1992.

[106] Morais, M.C. and Pacheco, A. (2006) Combined CUSUM-Shewhart Schemes for Binomial Data. *Economic Quality Control*, **21(1)**, 43-57, 2006.

[107] Moustakides,G.V. (1986) Optimal Stopping Times for Detecting Changes in Distribution. *The Annals of Statistics*, **14**, 1379-1387, 1986.

[108] Moustakides, G.V. (1998) Quickest Detection of Abrupt Changes for a Class of Random Processes. *IEEE Transactions on Information Theory*, **44**, 1965-1968, 1998.

[109] Moustakides, G.V. (2004) Optimality of the CUSUM Procedure in Continuous Time. *The Annals of Statistics*, **32(1)**, 302-315, 2004.

[110] Nally, R.M. and Hart, B.T. (1997) Use of CUSUM Methods for Water-Quality Monitoring in Storages. *Environmental Science and Technology*, **31(7)**, 2114-2119, 1997.

[111] Neftci, S.N. (2000) Value at Risk Calculations, Extreme Events, and Tail Estimation. *Journal of Derivatives*, 23-37, 2000.

[112] Page, E.S. (1954) Continuous inspection schemes. *Biometrika*, **41**, 100-115, 1954.

[113] Page, E.S. (1955) A Test for a Change in a Parameter Occurring at an Unknown Point. *Biometrika*, **41**, 100-115, 1955.

[114] Pan, J. and Chen,J. (2008) U-Statistic Based Modified Information Criterion for Change Point Problems. *Communications in Statistics-Theory and Methods*, **37(17)**, 2687-2712.

[115] Philips, K.T., Yashchin, E. and Stein, M.D. (2003) Using Statistical Process Control to Monitor Active Managers. *The Journal of Portfolio Management*, Fall 2003.

[116] Pollak, M. (1985) Optimal Detection of a Change in Distribution. *The Annals of Statistics*, **13(1)**, 206-227, 1985.

[117] Pollak, M. (1987) Average Run Lengths of an Optimal Method of Detecting a Change in Distribution. *The Annals of Statistics*, **15(2)**, 749-779, 1987.

[118] Ploberger, W., Krämer, W. and Alt, R. (1989) A Modification of the CUSUM Test in the Linear Regression Model with Lagged Dependent Variables. *Empirical Economics*, **2**, 65-75, 1989.

[119] Qiu, P. and Hawkins, D. (2001) A Rank Based Multivariate CUSUM Procedure. *Technometrics*, **43(2)**, 120-132, 2001.

[120] Reiss, R.D. and Thomas, M. (1997) *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhäuser Verlag, Basel.

[121] Reynolds, R.M. Jr. and Stoumbos,G.Z. (2010) Robust CUSUM Charts for Monitoring the Process Mean and Variance. *Quality and Reliability Engineering International*, **26**, 453-473, 2010.

[122] Ritov, Y. (1990) Decision Theoretic Optimality of the CUSUM Procedure. *The Annals of Statistics*, **18**, 1464-1469, 1990.

[123] Roberts, S.W. (1966) A Comaprison of Some Control Chart Procedures. *Technometrics*, **8**, 411-430, 1966.

[124] Robbins, M. W., Lund, R. B., Gallagher, C. M., Lu, Q. (2011) Changepoints in the North Atlantic Tropical Cyclone Record, *J. Amer. Statist. Assoc*, **106** (493), 89-99.

[125] Rochelle, E.W., Serryn, E., Bert, V., Graeme, W., and Aileen, J.P. (2008) Applying Cusum-Based Methods for the Detection of Outbreaks of Ross River Virus Disease in Western Australia. *BMC Medical Informatics and Decision Making*, **8(37)**, 2008.

[126] Rogerson, P. (2005) Monitoring Spatial Maxima. *Journal of Geographical Systems*, **7**, 101-114, 2005.

[127] SCHWARZ, G.E. (1978) Estimating the Dimension of a Model. *Annals of Statistics*, **6(2)**, 461-464.

[128] SEVERO,M. AND GAMA,J. (2010) Change Detection with Kalman Filter and CUSUM. *Lecture Notes in Computer Science*, **6202/2010**, 148-162, 2010.

[129] SHEWHART,W.A. (1931) *Economic Control of Quality of Manufactured Product*, Van Nostrand, Princeton.

[130] SHIAU, H.J. AND YA-CHEN HSU, Y.C. (2005) Robustness of the EWMA Control Chart to Non-normality for Autocorrelated Processes. *Quality Technology and Quantitative Management*, **2(2)**, 125-146, 2005.

[131] SRIVASTAVA, M.S. AND WORSLEY, K.J. (1986) Likelihood Ratio Tests for a Change in the Multivariate Normal Mean. *Journal of American Statistical Association*, **81(393)**, 199-204.

[132] SHU, L., YEUNG, H. AND JIANG, W. (2010) An Adaptive CUSUM Procedure for Signaling Process Variance Changes of Unknown Sizes. *Journal of Quality Technology*, **42**, 69-85, 2010.

[133] SIBBERTSEN, P. (2000) Robust CUSUM-M test in the Presence of Long-Memory Disturbances, 2000.

[134] SIEGMUND, D. (1988) Confidence Sets in Change Point Problem. *International Statistical Review*, **56**, 31-48.

[135] SMITH,J.A. (1987) Estimating the Upper Tail of Flood Frequency Distribution. *Water Resources Research*, **23(8)**, 1557-1666, 1987.

[136] STEINER,S.H., COOK,R.J. AND FAREWELL,V.T.(1999) Monitoring Paired Binary Surgical Outcomes Using Cumulative Sum Chart. *Statistics in Medicine*, **18(1)**, 69-86, 1999.

[137] STEINER,S.H., COOK,R.J. AND FAREWELL,V.T. (1999) Monitoring Paired Binary Surgical Outcomes Using Cumulative Sum Chart. *Statistics in Medicine*, **18(1)**, 69-86, 1999.

[138] STEWARDSON, D. AND COLEMAN, S. (2001) Using the Summed Rank Cusum for Monitoring Environmental Data from Industrial Processes. *Journal of Applied Statistics, Taylor and Francis Journals*, **28(3-4)**, 469-484, 2001.

[139] TIBSHIRANI, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, **58(1)**, 267-288.

[140] TSONIS, A.A., ELSNER, J.B., HUNT, A.G. AND JAGGER, T.H. (2005) Unfolding the Relation Between Global Temperature and ENSO. *Geophysical Research Letters*, doi: 10.1029/2005GL022875, 2005.

[141] VOSTRIKOVA, L.J. (1981) Detecting 'Disorder' in Multidimensional Random Processes, *Soviet Mathematics Doklady*, **24**, 55-59.

[142] WANG, A.L. (1990) Exponential Cusum Scheme for Detecting a Decrease in Mean. *The Indian Journal of Statistics, Series B*, **52(1)**, 105-114, 1990.

[143] Weiss, H.C. (2011) Detecting Mean Increases in Poisson INAR(1) Processes with EWMA Control Charts. *Journal of Applied Statistics*, **38(2)**, 383-398, 2011.

[144] Weiss, H.C. (2009) CUSUM Monitoring of First-Order Integer-Valued Autoregressive Processes of Poisson Counts. *Journal of Quality Technology*, **41(4)**, 389-400, 2009.

[145] Worsley, K.J. (1986) Confidence Regions and Tests for a Change-point in a Sequence of Exponential Family Random Variables. *Biometrika*, **73**, 91-104.

[146] Wu, Z., Jiao, J. and Liu, Y. A Poisson CUSUM Chart for Detecting Large Shifts in Fraction Non Conforming.

[147] Xia, Z., Guo, P., and Zhao, W. (2009) Monitoring Structural Changes in Generalized Linear Models. *Communications in Statistics-Theory and Methods*, **38(11)**, 1927-1947, 2009.

[148] Yang, T.Y. and Kuo, L. (2001) Baysian Binary Segmentation Procedure of a Poisson Process with Multiple Changepoints. *Journal of Computational and Graphical statistics*, **10(4)**, 772-785.

[149] Yao, Y.C. (1988) Estimating the Number Of Change Points Via Schwarz' Criterion. *Statistics and Probability Letters*, **6**, 181-189.

[150] Yashchin, E. (1993) Performance of CUSUM Control Schemes for Serially Correlated Observations. *Technometrics*, **35(1)**, 37-52, 1993.

[151] Yi, G., Coleman, S. and Ren, Q. (2006) CUSUM Method in Predicting Regime Shifts and Its Performance in Different Stock Markets Allowing for Transaction Fees. *Journal of Applied Statistics*, **33(7)**, 647-661, 2006.

[152] Zdansky, J. (2006) Speaker Change Detection via Binary Segmentation Technique and Informational Approach. *SPECOM'2006*.

[153] Zeileis, A. (2005) A Unified Approach to Structural Change Tests Based on ML Scores, F Statistics, and OLS Residuals. *Econometric Reviews*, **24(4)**, 445-466, 2005.

[154] Zeileis, A. (2005) Detecting Longitudinal Heterogeneity in Generalized Linear Models. *Adaptive Information Systems and Modelling In Economics and Management Science, Interdisciplinary Studies in Economics and Management*, **5(4)**, 253-259, 2005.

[155] Zeileis, A. (2006) Implementing a Class of Structural Change Tests: An Econometric Computing Approach. *Computational Statistics and Data Analysis*, **50**, 2987-3008, 2006.

[156] Zeileis, A., Kleiber, C. and Krämer, W. (2003) Testing and Dating of Structural Changes in Practice. *Computational Statistics and Data Analysis*, **44**, 109-123, 2003.

[157] Zhang, N. R. and Siegmund, D. O. (2007) A Modified Bayes Information Criterion with Applications to the Analysis of Comparative Genomic Hybridization Data. *Biometrics*, **63**, 22-32.

[158] Shen, J.J. and Zhang, N.R. (2011) Change-Point Model on Non-Homogeneous Poisson Processes with Application in Copy Number Profiling by Next-Generation DNA Sequencing. Preprint.

[159] Zou, H. and Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B*, **67(2)**, 301-320.